Numero 296
Aprile 1993
Anno XXVI
Volume L

LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

Scarsità di risorse e degrado come causa di violenti conflitti

Uno studio condotto da ricercatori canadesi e statunitensi dimostra come crescita demografica e deterioramento ambientale siano alla base dell'insorgere di devastanti guerre civili e di accesi conflitti fra popoli

di Thomas F. Homer-Dixon, Jeffrey H. Boutwell e George W. Rathjens

Entro 50 anni è probabile che la popolazione umana supererà i nove miliardi di unità, mentre la produzione economica globale sarà cinque volte superiore a quella attuale. Queste due tendenze saranno forse le principali responsabili di una rapida diminuzione della disponibilità di risorse rinnovabili. Calerà l'area totale destinata all'agricoltura ad alta produttività, così come l'estensione delle foreste e il numero delle specie animali che vi vivono. Le generazioni future assisteranno anche al progressivo esaurimento e degrado di falde acquifere, fiumi e altri corpi idrici, al declino dell'attività di pesca, all'ulteriore riduzione dell'ozono stratosferico e a una sensibile alterazione climatica.

L'aggravarsi di questi problemi ambientali può facilmente portare a guerre civili o conflitti internazionali. Da decenni, gli scienziati interessati a questi problemi mettono in guardia da questo pericolo, ma il dibattito è stato frenato dalla mancanza di dati attendibili. Per colmare questa lacuna, abbiamo creato un gruppo di 30 ricercatori con il compito di analizzare un insieme di casi specifici. In studi commissionati dall'Università di Toronto e dall'American Academy of Arts and Sciences, questi esperti hanno reso noti i primi risultati.

I dati raccolti portano verso una conclusione preoccupante: la scarsità di risorse rinnovabili è già tra le cause di violenti conflitti in molte zone del mondo in via di sviluppo. Questi conflitti prefigurano forse l'insorgere di violenze analoghe nei prossimi decenni, particolarmente in quei paesi poveri in cui la disponibilità sempre minore di acqua, di foreste e soprattutto di terreni fertili, insieme con la rapida crescita demografica, è già all'origine di gravi difficoltà.

Prima di passare all'esame dei risultati dell'indagine, va sottolineato come l'ambiente sia solo una variabile di una serie di fattori politici, economici e sociali in grado di generare disordini. Anzi, secondo alcuni, la scarsità di risorse rinnovabili è solo una variabile secondaria che occasionalmente collega preesistenti fattori politici ed economici al conflitto sociale conseguente.

I dati che abbiamo raccolto conducono a una prospettiva molto diversa. La scarsità delle risorse può essere un fattore determinante per cambiarne la gestione politica ed economica e indurre chi detiene il potere a volgere ulteriormente a proprio vantaggio una iniqua distribuzione delle stesse. Inoltre, la vulnerabilità dell'ecosistema contribuisce spesso in modo significativo alla riduzione delle risorse rinnovabili. Questa vulnerabilità è in parte un dato fisico: nelle regioni tropicali, per esempio, lo spessore del suolo nelle zone montane non dipende dalle istituzioni sociali o dal comportamento umano. E infine, in molte regioni del globo, il degrado ambientale sembra aver superato il punto di non ritorno. In questi casi, anche un intervento che rimuova le cause politiche, economiche e culturali del degrado può continuare a contribuire alla disgregazione sociale. In altri termini, una volta irreversibile, il degrado ambientale diventa una variabile indipendente.

Spesso gli scettici seguono un'altra linea di ragionamento e affermano che i conflitti dovuti alla carenza di risorse sono di scarso interesse, in quanto sono sempre esistiti nella storia dell'uomo. Noi sosteniamo, invece, che nei prossimi 50 anni si verificherà un depauperamento delle risorse naturali di rapidità, complessità e dimensioni tali da non aver precedenti nella storia. In pochi decenni si potrebbe arrivare alla deforestazione di interi paesi; nell'arco di una generazione potrebbe scomparire buona parte del terreno agricolo di una regione, e nel giro di soli 20 anni si potrebbe avere una grave diminuzione dell'ozono.

A differenza delle risorse non rinnovabili - come i combustibili fossili e i minerali ferrosi - quelle rinnovabili sono collegate in sistemi interdipendenti di elevata complessità, sulla base di relazioni non lineari e retroattive. L'eccessivo prelievo di una risorsa può portare, quando il sistema supera certe soglie critiche, a molteplici problemi ambientali imprevisti e all'improvviso manifestarsi di scarsità di risorse di altro tipo.

In base alle nostre ricerche, riteniamo che i disordini sociali e politici provocati dalle mutate condizioni ambientali non si configureranno come quelli che tradizionalmente siamo soliti catalogare come conflitti dovuti a penuria di risorse. Spesso, in passato, gruppi o nazioni hanno cercato di accaparrarsi risorse altrui. Durante la seconda guerra mondiale, per esempio, i giapponesi tentarono di assicurarsi petrolio, minerali e altre materie prime in Cina e nel Sud-Est asiatico.

Attualmente, però, molte delle risorse rinnovabili minacciate sono beni comuni - l'atmosfera e gli oceani, per esempio - ed è quindi improbabile che possano essere causa di scontri diretti. Inoltre, siamo arrivati a capire che la scarsità di risorse produce spesso effetti sociali insidiosi e cumulativi, come gli spostamenti di popolazioni e la rovina economica. Questi eventi, a loro volta, portano a conflitti tra gruppi etnici così come a scontri o sommosse civili. Anche se questi conflitti possono avere, sulla scarsità di risorse, effetti meno drammatici di una guerra, è probabile che si ripercuotano sulla sicurezza del mondo sviluppato e di quello in via di sviluppo.

Le azioni umane influiscono sulla disponibilità delle risorse rinnovabili soprattutto in tre modi. In primo luogo, i popoli possono ridurne la quantità o degradarne la qualità più velocemente della loro capacità di rigenerazione. Spesso il fenomeno viene indicato come consumo del «capitale-risorsa»: il capitale genera «reddito» cui si può attingere per il consumo umano. Un'economia sostenibile, quindi, si può definire come un'economia che lascia intatto il capitale, in modo che le generazioni future possano godere di redditi immutati. Co-

Una pattuglia militare in un villaggio dell'Assam, in India, dove nel 1983 la popolazione locale attaccò gli immigrati musulmani provenienti dal Bangladesh. Molti residenti accusavano da tempo gli immigrati di essersi appropriati delle migliori terre agricole della regione. Prima che arrivasse l'esercito a sedare i tumulti furono massacrati quasi 1700 bengalesi.

sì, se la rigenerazione del suolo in una regione agricola è di 0,25 millimetri all'anno, la perdita media non dovrebbe superare quella quantità.

La seconda causa di scarsità di risorse è la crescita demografica: può accadere, per esempio, che una certa quantità d'acqua vada divisa tra un numero crescente di persone. L'ultima causa è il cambiamento nella distribuzione di una risorsa all'interno di una società. Questa variazione può condurre alla concentrazione delle risorse nelle mani di pochi, assoggettando il resto della popolazione a condizioni di estrema ristrettezza.

Questi tre fattori possono operare singolarmente oppure in combinazione. In alcuni casi, basta la crescita demografica per ingenerare tensioni sociali. Il Bangladesh, per esempio, non soffre di degrado del suolo o di erosione delle terre agricole: le inondazioni annuali del Gange e del Brahmaputra depositano uno strato di limo che conserva la fertilità delle vaste pianure alluvionali del paese.

Secondo le Nazioni Unite, però, gli attuali 120 milioni di abitanti del Bangladesh arriveranno a 235 milioni nel 2025. Con circa 0,08 ettari pro capite, le terre coltivate sono già adesso disperatamente insufficienti. La densità della popolazione è di 785 persone per chilometro quadrato (per confronto, la densità del vicino stato indiano dell'Assam è di 284 persone per chilometro quadrato). Dato che tutto il terreno fertile è stato sfruttato, la crescita demografica dimezzerà entro il 2025 la disponibilità pro capite di terre coltivabili. Le inondazioni e l'inadeguatezza delle istituzioni nazionali e locali preposte al controllo delle acque acuiscono il problema della mancanza di terra e la spaventosa povertà e i disordini che ne conseguono.

Negli ultimi 40 anni, milioni di persone sono emigrate dal Bangladesh verso zone limitrofe dell'India, dove il tenore di vita è spesso migliore. I dati dettagliati sugli spostamenti sono scarsi: il Governo del Bangladesh è restio ad ammettere l'entità di questa migrazione perché il fenomeno è diventato fonte di attriti con l'India. Ciononostante, uno dei nostri ricercatori, Sanjoy Hazarika, inviato del «New York Times» a Nuova Delhi, ha raccolto stime e dati demografici attendibili e ne ha concluso che gli emigranti del Bangladesh e i loro discendenti hanno fatto aumentare di 15 milioni la popolazione delle zone indiane confinanti. (Solo uno o due milioni di questi possono essere attribuiti alle migrazioni avvenute nel 1971 durante la guerra tra India e Pakistan che portò alla nascita del Bangladesh.)

Questo enorme flusso migratorio ha prodotto grandi cambiamenti sociali negli Stati indiani ospitanti. Conflitti sono sorti dalla diversa distribuzione delle terre e da modifiche negli equilibri politici ed economici tra gruppi religiosi ed etnici. Per esempio, i lalung dell'Assam hanno a lungo opposto resistenza agli immigrati musulmani bengalesi, accusati di appropriarsi dei migliori appezzamenti agricoli della zona. Agli inizi del 1983, durante l'elezione aspramente contestata ad alcune cariche federali dello Stato, la violenza finì con l'esplodere: nel villaggio di Nellie, i lalung massacrarono circa 1700 bengalesi in un tumulto durato cinque ore.

Nello Stato di Tripura, le popolazioni originarie di fede buddista e cristiana sono ora meno del 30 per cento del totale. La percentuale restante è formata da immigrati indù provenienti dal Pakistan orientale o dal Bangladesh. Questo spostamento nell'equilibrio etnico provocò tra il 1980 e il 1988 una violenta ribellione che venne sedata solo dopo la promessa da parte del Governo di restituire le terre agli abitanti di Tripura e di porre fine all'emigrazione dal Bangladesh. Questo accordo, però, è in pericolo perché l'emigrazione è proseguita.

Naturalmente, gli spostamenti di popolazioni in questa parte del Sud-Est asiatico non sono nuovi. Durante il periodo coloniale, gli inglesi fecero affluire indù da Calcutta per amministrare l'Assam, e il bengalese divenne la lingua ufficiale. Gli assamesi sono quindi molto sensibili a una perdita di controllo politico e culturale nel loro Stato. Il Governo indiano ha spesso incoraggiato l'immigrazione per raccogliere voti. Oggi, però, è chiaro che all'esodo dal Bangladesh contribuisce la maggiore densità demografica. Per quanto importanti siano i fattori religiosi e politici, è innegabile che nel Bangladesh la penuria di terra coltivabile sia una spinta al conflitto.

In altre aree del mondo, tutte e tre le cause di scarsità concorrono a produrre disordini. La crescita demografica e la riduzione in qualità e quantità delle risorse rinnovabili possono portare a progetti di sviluppo di grande scala che possono alterare l'accesso alle risorse. Questo processo può portare a una minore disponibilità di risorse per i gruppi più poveri, alle cui esigenze si oppongono violentemente le *élite* al potere. Ne è un esempio la disputa iniziata nel 1989 tra Mauritania e Senegal nella valle del fiume Senegal, che segna il confine tra questi due paesi.

Il Senegal dispone di una certa abbondanza di terre agricole, molte delle quali, però, soffrono di un'intensa erosione eolica, di aumento di salinità per eccesso di irrigazione, di perdita di elementi nutritivi, e di compattazione dovuta a pratiche agricole intensive. Il paese ha una densità di 38 abitanti per chilometro quadrato e un incremento demografico annuo del 2,7 per cento: in 25 anni la popolazione può raddoppiare. Al contrario, se si fa eccezione per la valle del fiume Senegal lungo il confine meridionale e per alcune oasi, la Mauritania è per lo più desertica o coperta di praterie semiaride. La sua densità di popolazione è molto bassa, circa 2 abitanti per chilometro quadrato, e l'incremento demografico annuo è del 2,8 per cento. La FAO (Food and Agricultural Organization) delle Nazioni Unite ha incluso Mauritania e Senegal nell'elenco dei paesi privi di terreni coltivabili sufficienti per sostentare la popolazione attuale o quella prevista a meno che non vi sia un forte aumento dei fattori produttivi, come fertilizzanti o nuove tecniche di irrigazione.

Normalmente, le vaste piane alluvionali che costeggiano il fiume Senegal consentono produttive attività di coltivazione, allevamento e pesca basate sulle piene annuali del fiume. Negli anni settanta, però, la prospettiva di una cronica penuria di derrate alimentari e una grave siccità spinsero le autorità della regione a chiedere finanziamenti internazionali per la diga di Manantali sul fiume Bafing, affluente del Senegal nel Mali, e per lo sbarramento contro l'intrusione salina presso la foce del fiume Senegal tra il Senegal e la Mauritania. Le dighe erano progettate per regolare la portata del fiume in modo da fornire energia idroelettrica, espandere l'agricoltura irrigata e aumentare la disponibilità idrica nella stagione secca, permettendo trasporti su chiatta per tutto l'anno dall'Oceano Atlantico al Mali, paese senza sbocchi sul mare.

Ma il progetto ebbe conseguenze sfortunate e impreviste. Come ha dimostrato l'antropologo Michael M. Horowitz, della State University of New York a Binghamton, la costruzione delle nuove dighe fece aumentare il valore delle terre lungo il fiume dove stava per realizzarsi un'agricoltura intensiva. La classe dominante mauritana, formata in prevalenza da mauri bianchi, riscrisse la legislazione relativa alla proprietà terriera, abrogando di fatto il diritto dei mauri neri a praticare agricoltura, allevamento e pesca lungo le rive mauritane del fiume.

C'è una lunga storia di discriminazioni razziali praticate dai mauri bianchi nei confronti dei loro compatrioti neri non arabi. Nella primavera del 1989, l'uccisione di alcuni coltivatori senegalesi da parte di mauri nel bacino del fiume scatenò un'ondata di violenza etnica tra i due paesi. In Senegal, quasi tutti i 17 000 negozi di proprietà dei mauri vennero distrutti e i loro proprietari espulsi in Mauritania. In entrambi i paesi, diverse centinaia di persone vennero uccise, e le due nazioni furono a un passo dalla guerra. Il regime mauritano sfruttò l'occasione per introdurre una nuova legislazione che dichiarava «senegalesi» i mauri neri che vivevano lungo il fiume; li privava quindi della cittadinanza e ne confiscava le proprietà. Ben 70 000 mauri neri furono espulsi con la forza in Senegal, da dove alcuni iniziarono a compiere scorrerie per riappropriarsi del bestiame confiscato. Attualmente, i due paesi hanno riallacciato le relazioni diplomatiche, ma nessuno dei due ha consentito il rientro della popolazione espulsa o l'ha risarcita per le perdite subite.

DISPONIBILITÀ DI TERRENO AGRICOLO

1989 | 2025

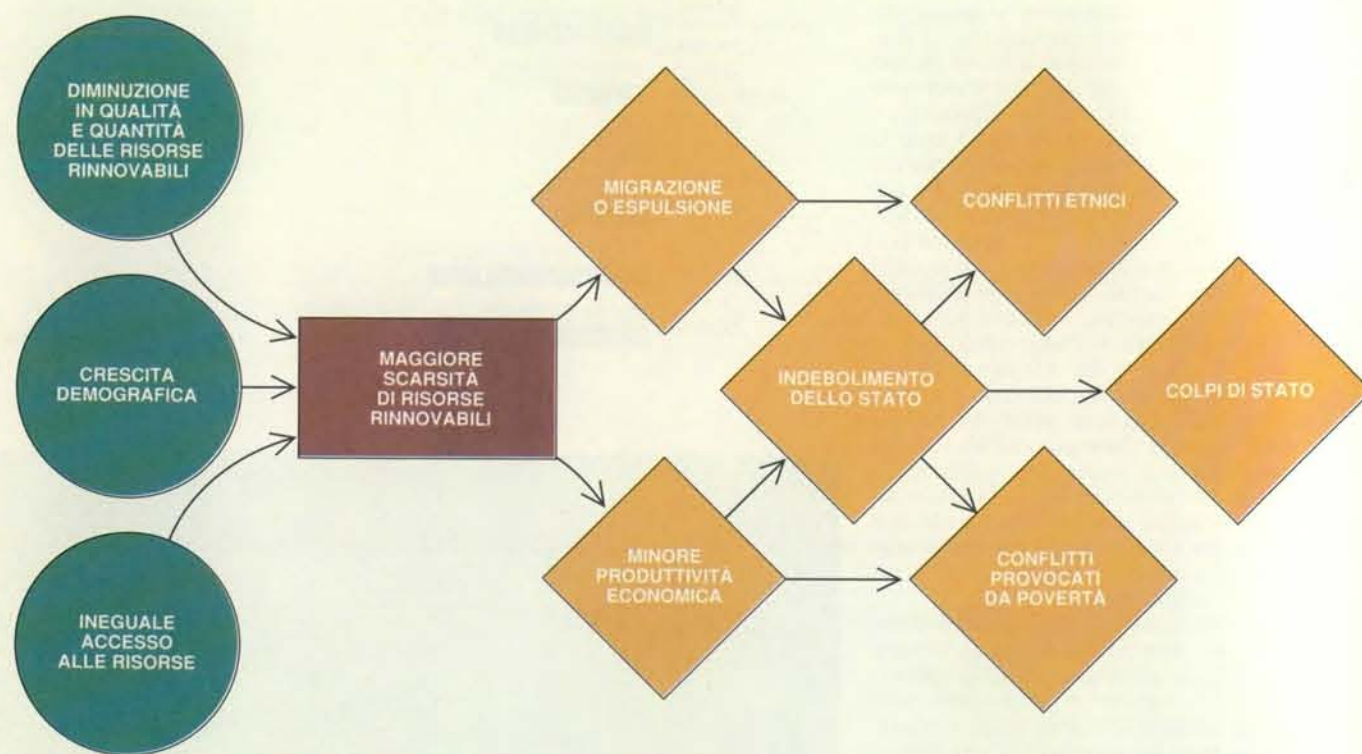
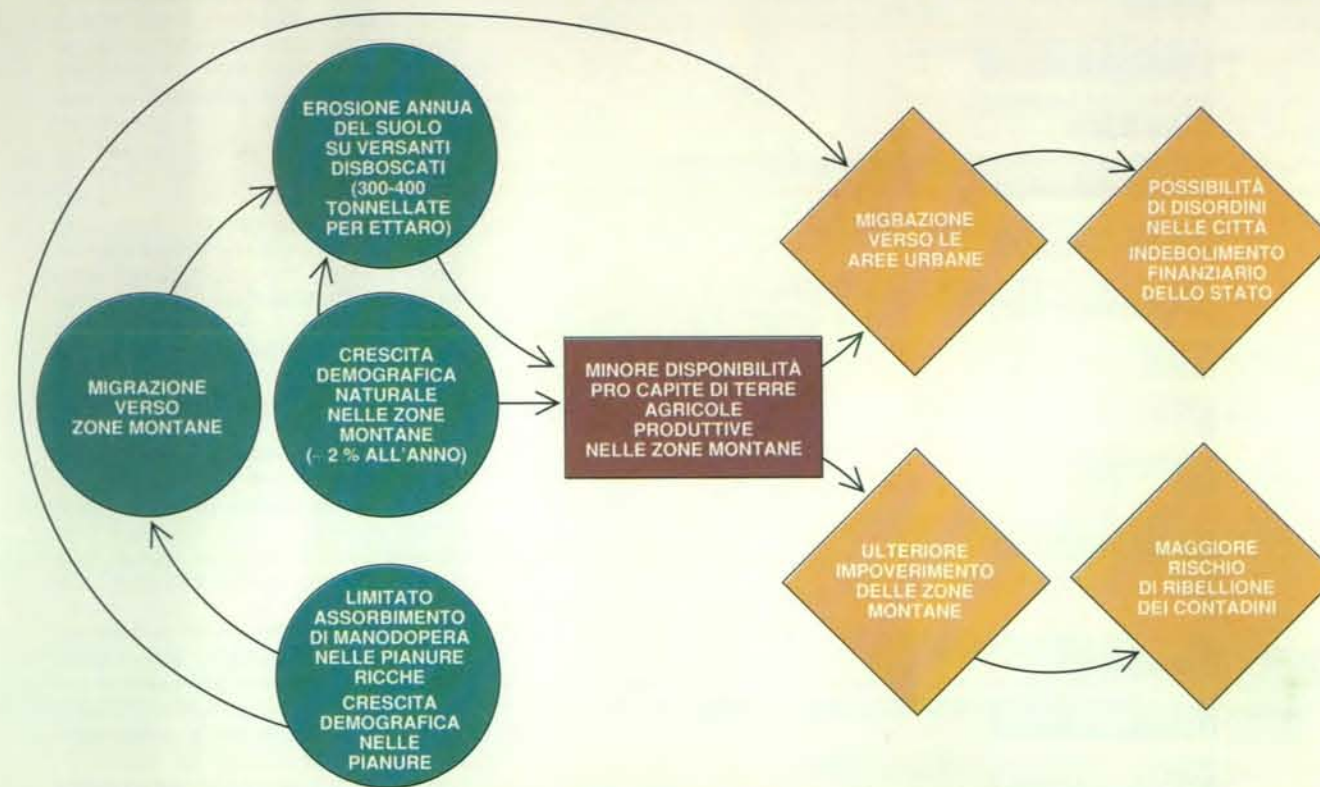
AFRICA
AMERICA SETTENTRIONALE E CENTRALE
AMERICA MERIDIONALE
ASIA
EUROPA
EX UNIONE SOVIETICA
OCEANIA

0,0 | 0,5 | 1,0 | 1,5 | 2,0

SUPERFICIE COLTIVATA PER PERSONA (ETTARI)

FONTE: World Resources Institute

Si ritiene che da ora al 2025 la disponibilità di aree per l'agricoltura subirà una diminuzione (*in alto*) sia per la crescita demografica sia per il degrado delle terre coltivabili. Nelle Filippine la mancanza di terreno fertile ha spinto gli agricoltori poveri sui versanti montani (*in basso*), la cui coltivazione senza opere di terrazzamento provoca una grave erosione, come si vede dagli squarci color terra sui pendii.

In molte parti del mondo assistiamo a un processo causale un po' differente: l'accesso ineguale alle risorse si combina con la crescita demografica per produrre danni ambientali. Questo fenomeno può contribuire all'impoverimento economico, causa di disordini e ribellioni. Nelle Filippine, la politica coloniale spagnola e statunitense ha lasciato una distribuzione delle terre fortemente iniqua. A partire dagli anni sessanta, l'introduzione nelle pianure di tecnologie agricole innovative ha consentito un netto aumento nella produzione di cereali

sia per il consumo interno sia per il mercato internazionale cosicché è stato possibile ridurre l'enorme debito con l'estero contratto dal paese.

Questa modernizzazione ha fatto crescere la domanda di manodopera agricola. Sfortunatamente, però, il miglioramento è stato vanificato da un incremento demografico oscillante tra il 2,5 e il 3,0 per cento. Insieme con l'iniqua distribuzione delle terre coltivabili e con la crisi economica della prima metà degli anni ottanta, l'aumento della popolazione ha prodotto un'ondata di disoccupazione agricola.

In assenza di una sufficiente industrializzazione rurale o urbana per assorbire questo eccesso di manodopera, si è verificata un'inevitabile riduzione dei salari. In condizioni economiche disperate, milioni di coltivatori poveri e di braccianti senza terra sono emigrati nelle bidonville di città già sovraffollate, come Manila; milioni di altri si sono spostati in territori meno produttivi - e spesso ecologicamente molto vulnerabili - come i versanti delle zone montane.

In queste aree montane, gli abitanti usano il fuoco per aprire radure in zone precedentemente a foresta o già disboscate. Molto scarsa è la loro capacità di proteggere il fragile ecosistema. Il disboscamento su piccola scala, la produzione di carbone vegetale e l'agricoltura fondata sul debbio provocano spesso erosione, smottamenti e modificazioni nell'equilibrio idrogeologico. Tutto questo ha creato un circolo vizioso di bassa produzione alimentare, disboscamento di nuovi siti e ulteriore degrado del territorio. In molte zone sta diventando difficile anche trovare terre solo marginalmente produttive, e le condizioni economiche sono critiche per i contadini.

Il paese è stato turbato da gravi disordini civili per molti decenni, ma due ricercatori, Celso R. Roque, già Sottosegretario all'ambiente delle Filippine, e la sua collega Maria I. Garcia, sostengono che tra le cause dell'attuale insurrezione di stampo comunista stia assumendo sempre maggiore importanza la scarsità di risorse. La guerriglia nei territori montani - che vede anche assalti alle basi militari - trae origine dalla povertà dei braccianti agricoli e dei piccoli proprietari terrieri di quelle regioni che sfuggono in larga misura al controllo del Governo centrale. Negli anni settanta e ottanta, il New People's Army e il National Democratic Front hanno trovato tra i contadini delle montagne il terreno adatto per la diffusione di un'ideologia rivoluzionaria, soprattutto là dove proprietari terrieri e amministrazioni locali particolarmente repressivi non hanno lasciato alternative se non la ribellione o la morte per fame. I rivoluzionari hanno fatto leva sulle credenze indigene e sulle strutture sociali per aiutare i contadini a dare uno sbocco al loro malcontento.

Concatenazioni di cause simili a quelle occorse nelle Filippine si riscontrano in molte altre regioni del pianeta, tra cui quella himalayana, il Sahel, l'Indonesia, il Brasile e la Costa Rica. La crescita demografica e l'iniqua distribuzione delle terre coltivabili spingono un numero enorme di persone verso le città o verso terre marginali. In queste ultime, i nuovi arrivati provocano danni ambientali e la loro povertà si cronicizza. Le masse diseredate, infine, possono divenire terreno di ribellioni endemiche o, con nuove migrazioni, innescare altrove conflitti etnici o disordini urbani.

Una simile combinazione di fattori si trova nella breve ma devastante «Guerra del pallone» scoppiata nel 1969 tra El Salvador e Honduras, così detta per aver avuto origine da scontri verificatisi a Tegucigalpa durante una partita di calcio fra le squadre dei due paesi. Come ha dimostrato William H. Durham, della Stanford University, le trasformazioni avvenute nell'agricoltura e nella distribuzione terriera a partire dalla metà del XIX secolo portarono a una concentrazione di agricoltori poveri nelle regioni montuose di El Salvador. Questi contadini dimostrarono una certa capacità di conservazione delle terre, ma la crescita demografica produsse deforestazione ed erosione sui versanti più scoscesi. Un incremento demografico annuo del 3,5 per cento ridusse ulteriormente la disponibilità di terreno coltivabile e di conseguenza molti si trasferirono nel vicino Honduras. La loro successiva espulsione dall'Honduras provocò lo scoppio di una guerra nel corso della quale, in pochi giorni, furono uccise diverse migliaia di persone. Durham osserva che la competizione per la terra, all'origine di questo conflitto, non venne risolta nel dopoguerra e contribuì in larga misura alla successiva guerra civile che per un decennio insanguinò la regione.

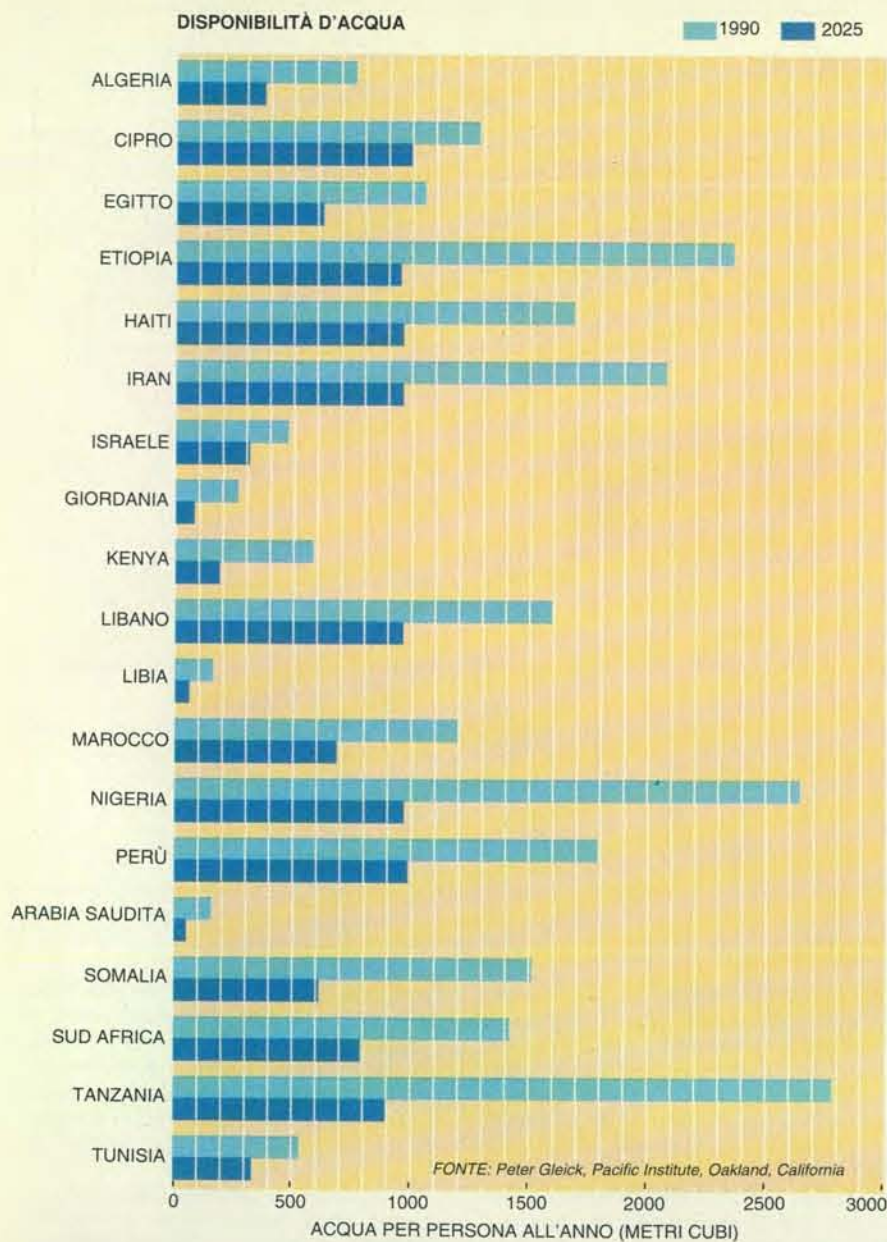
In Sud Africa, la politica di apartheid adottata nel passato dal regime dei bianchi ha concentrato milioni di neri nei territori meno produttivi ed ecologicamente più vulnerabili. L'elevata natalità ha fatto aumentare eccessivamente la densità di popolazione. Nel 1980, le aree rurali del territorio (*homeland*) del Ciskei ospitavano 82 persone per chilometro quadrato, mentre la vicina Provincia del Capo aveva una densità rurale di due abitanti per chilometro quadrato. Gli abitanti del Ciskei avevano, e hanno, una ridotta disponibilità finanziaria e scarse capacità di gestione delle risorse: sono quindi facili vittime delle soperchierie di amministratori locali corrotti.

In una situazione di questo genere, uno sviluppo sostenibile è impossibile. Vaste aree sono state completamente spogliate degli alberi per ottenere combustibile, ridotte a terra nuda e polverosa dall'eccessivo sfruttamento a pascolo e private dello strato fertile. Una relazione del 1980 concludeva che circa il 50 per cento del territorio del Ciskei era parzialmente o gravemente eroso e quasi il 40 per cento dei suoi pascoli era esaurito. Questa perdita di risorse, insieme con la mancanza di occupazioni alternative e al trauma sociale dell'apartheid, ha generato, nei territori autonomi, una crisi di sopravvivenza, in seguito alla quale migliaia di persone sono emigrate nelle città sudafricane. Si è avuta così una rapida crescita di insediamenti abusivi e la formazione di comunità illegali che sono teatro di continui scontri e minacciano di compromettere la transizione del paese alla stabilità democratica.

La diminuzione delle risorse naturali può indebolire la capacità e l'autorità amministrativa di un paese, con la possibile conseguenza di violenti attacchi allo Stato da parte di oppositori politici e militari. Alimentando la povertà rurale e il fenomeno dell'inurbamento, la scarsità di risorse rinnovabili fa crescere il numero di persone bisognose di assistenza pubblica. Per far fronte all'aumento della popolazione urbana, spesso si introducono sussidi che distorcono i prezzi e impediscono una corretta destinazione degli investimenti.

Allo stesso tempo, la perdita di risorse rinnovabili può ridurre la produzione di beni, con una conseguente diminuzione delle entrate fiscali. In alcuni paesi, il crescente divario fra domanda e risorse economiche può acuire il risentimento popolare, erodere la legittimità del potere amministrativo e inasprire la competizione tra le élite in lotta per salvaguardare i propri privilegi.

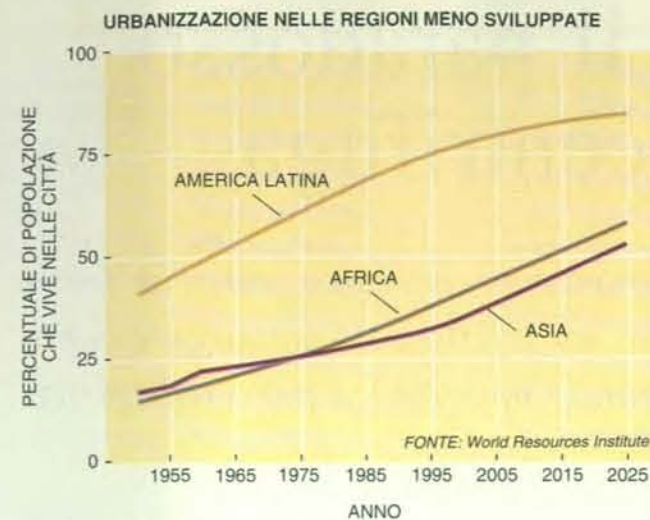
Il taglio degli alberi per destinarli ai mercati esteri, come avviene nel Sud-Est asiatico e in Africa occidentale, produce sul breve termine un guadagno per alcune élite e può alleviare il debito con l'estero. Sul lungo periodo, però, mette in pericolo la produttività. L'eliminazione delle foreste riduce la capacità del terreno di trattenere l'acqua durante i periodi di intense precipitazioni. Piene improvvise danneggiano strade, ponti, sistemi d'irrigazione e altre infrastrutture di valore. Il materiale eroso dei versanti montani ostruisce i fiumi, riducendone la navigabilità e la capacità di produrre energia idroelettrica. La deforestazione può anche danneggiare la produzione agricola alterando i cicli idrologici regionali e provocando l'interramento dei bacini artificiali e dei canali di irrigazione (si veda l'articolo *La contabilità delle risorse ambientali* di Robert Repetto in «Le Scienze» n. 288, agosto 1992).

Considerando la situazione cinese, Václav Smil dell'Università del Manitoba ha fatto una valutazione dell'effetto combinato dei problemi ambientali sulla

È probabile che la mancanza d'acqua diventi sempre più grave. Nel 2025, diverse nazioni (*in alto*) avranno meno di 1000 metri cubi d'acqua per persona, la quantità minima prevista in un paese industrializzato. In Etiopia l'acqua è già così scarsa che alcune donne, per rifornirsene, sono costrette a tragitti di chilometri (*in basso*).

produttività. I principali aggravi economici identificati da Smil sono la riduzione dei raccolti a causa dell'inquinamento delle acque, del suolo e dell'aria; la maggiore morbilità umana dovuta all'inquinamento atmosferico; la riduzione degli appezzamenti coltivabili a causa degli interventi edilizi e dell'erosione; la perdita di elementi nutritivi causata dall'erosione e dalla deforestazione; e la diminuzione di legname in seguito a tecniche antiquate di raccolta. Smil stima che l'attuale costo annuo di tutto ciò ammonti a circa il 15 per cento del prodotto interno lordo della Cina, ed è convinto che sia destinato a crescere rapidamente nei prossimi decenni. Stima anche che dieci milioni di cinesi cercheranno di abbandonare le regioni più povere del paese, quelle interne e settentrionali dove acqua e legna da ardere sono disperatamente scarse e il territorio è spesso gravemente compromesso, per raggiungere le città costiere in piena crescita. Prevede aspre contese tra queste regioni sulla ripartizione delle risorse idriche e sulla migrazione. Nel loro insieme, queste tensioni economiche e politiche possono notevolmente indebolire lo Stato cinese.

In Medio Oriente la mancanza di acqua sarà sempre più sensibile e contribuirà forse ad aggravare le tensioni politiche. Le stime variano, ma Miriam R. Lowi della Princeton University valuta in 1950 milioni di metri cubi (Mm^3) la disponibilità media annua di acqua dolce per Israele. Il 60 per cento è acqua presente nel sottosuolo; il resto viene dai fiumi, dalle piene e dal riciclaggio di acque reflue. L'attuale domanda di Israele - compresa quella degli insediamenti nei territori occupati e nelle alture del Golan - è di circa 2200 Mm^3. Al deficit annuo di oltre 200 Mm^3 si fa fronte con un sovrasfruttamento delle falde acquifere.

In alcune zone di Israele e della Cisgiordania la falda freatica si è notevolmente abbassata, e questo impoverimento può far aumentare il grado di salinità nei pozzi e l'infiltrazione di acqua marina dal Mediterraneo. Allo stesso tempo, si prevede che nell'anno 2020 la popolazione israeliana passerà dagli attuali 4,6 milioni a 6,5 milioni, una stima che non tiene conto dell'immigrazione dall'ex Unione Sovietica. Sulla base di questa proiezione, la domanda d'acqua del paese potrebbe superare nel 2020 i 2600 Mm^3.

Delle tre principali falde acquifere da cui Israele dipende, due si estendono prevalentemente in Cisgiordania, anche se le loro acque defluiscono verso Israele. Circa il 40 per cento delle acque freatiche impiegate da Israele proviene quindi dai territori occupati. Per proteggere questa importante fonte di approvvigionamento il Governo israeliano ha imposto rigide limitazioni al consumo d'acqua in Cisgiordania. Dei 650 Mm^3 complessivi di acqua disponibili in quel territorio, agli arabi è consentito usarne solo 125 Mm^3. Il Governo israeliano limita

L'incontrollata crescita delle città, dovuta in parte all'acuirsi della povertà nelle campagne e in parte ai movimenti migratori, diverrà drammatica nel mondo in via di sviluppo (*a sinistra*). A Manila, l'insediamento abusivo chiamato «Montagne fumanti» è il luogo in cui finiscono i contadini poveri che arrivano per mare dalle zone di provincia (*a destra*). I filippini hanno dato questo nome all'insediamento per le continue esalazioni della discarica sulla quale è stato costruito.

il numero di pozzi che gli arabi possono perforare, la quantità di acqua che possono estrarre e i periodi in cui possono utilizzarla per l'irrigazione.

In Cisgiordania la discriminazione nell'accesso all'acqua è marcata: il consumo pro capite dei coloni ebrei è di circa quattro volte superiore a quello degli arabi. A questi ultimi è vietato scavare nuovi pozzi per usi agricoli, mentre la Mekorot (l'azienda idrica israeliana) ne ha scavati più di 30 per i coloni. L'agricoltura araba della regione ha risentito dell'aumento di salinità in alcuni pozzi dovuto allo scavo di pozzi israeliani più profondi nelle vicinanze. La politica idrica di Israele, insieme con la confisca di terre agricole a favore dei propri coloni e altre restrizioni imposte all'agricoltura palestinese, ha favorito l'abbandono delle campagne da parte di molti arabi cisgiordani, che ora sono disoccupati o avventizi.

Tutto il Medio Oriente si trova a fronteggiare una scarsità d'acqua che crea un complesso di problemi sempre più gravi, e secondo molti esperti questi problemi influiranno negativamente sulla stabilità della regione. Il problema dell'accesso all'acqua fu tra le cause delle tensioni che precedettero la guerra arabo-israeliana del 1967, quella guerra che diede a Israele il controllo su gran parte delle risorse idriche del bacino del Giordano. Nell'ambito degli attuali colloqui di pace sul Medio Oriente, sono in calendario numerosi incontri plurilaterali che hanno per tema i diritti sulle acque e testimoniano la preoccupazione suscitata dalla minaccia di penuria idrica.

Anche se future «guerre per l'acqua» sono possibili, sembrano improbabili data la superiorità militare di Israele. Più probabile è che, nel contesto degli storici contrasti etnici e politici, la mancanza d'acqua finisca con l'aggravare tensioni e agitazioni all'interno delle comunità che vivono nel bacino del Giordano, tesi ribadita anche, in una recente allocuzione tenuta al Congresso degli Stati Uniti, da Thomas Naff dell'Università della Pennsylvania.

È chiaro che la scarsità di risorse rinnovabili può favorire i conflitti, e che la frequenza di questi scontri andrà probabilmente aumentando. Ciononostante alcuni osservatori sostengono che la penuria di risorse non è importante di per sé. Ciò che importa, a loro giudizio, è la misura in cui essa danneggia una popolazione. La sofferenza umana potrebbe essere evitata se i sistemi politici ed economici fornissero gli incentivi e i mezzi per permettere ai popoli di alleviare gli effetti dannosi dei problemi ambientali.

La nostra ricerca non ha prodotto dati certi pro o contro questa argomentazione. È necessario ampliare le nostre conoscenze sulle variabili che influenzano la capacità umana di rispondere in modo innovativo alle modificazioni ambientali. È necessaria ingegnosità tecnica per elaborare, per esempio, nuove tecnologie agricole e forestali capaci di compensare il degrado dell'ambiente. È necessaria ingegnosità sociale per creare istituzioni che proteggano la popolazione dagli effetti del degrado e forniscano i giusti incentivi per l'innovazione tecnologica.

Spesso viene trascurato il ruolo che l'ingegnosità sociale può avere in rapporto all'ingegnosità tecnica. Un sistema complesso e stabile di mercati, di legislazioni, di istituti finanziari, scolastici e di ricerca è un prerequisito per l'elaborazione e la diffusione di numerose tecnologie - per esempio nuove sementi adatte a climi aridi e terreni erosi, tecnologie alternative di cottura che compensino la mancanza di legna da ardere e tecnologie per la conservazione delle acque. Non solo i paesi poveri sono carenti di queste risorse sociali, ma la loro capacità di crearle e conservarle sarà indebolita proprio da quelle difficoltà ambientali a cui in questi paesi si cerca di far fronte.

I dati che abbiamo esposto fanno ritenere che ci siano importanti legami causali tra scarsità di risorse rinnovabili e violenza. Per impedire il proliferare di tensioni e disordini, le nazioni dovrebbero puntare con maggior vigore alla conservazione delle risorse. Ciò significa che paesi ricchi e paesi poveri devono cooperare per ridurre la crescita demografica, favorire una più equa distribuzione interna e internazionale della ricchezza, e operare in direzione di uno sviluppo sostenibile.

BIBLIOGRAFIA

DURHAM WILLIAM H., *Scarcity and Survival in Central America: Ecological Origins of the Soccer War*, Stanford University Press, 1979.

GURR TED, *On the Political Consequences of Scarcity and Economic Decline* in «The International Studies Quarterly», 29, n. 1, marzo 1985.

LEONARD JEFFREY H. (a cura), *Environment and the Poor: Development Strategies for a Common Agenda*, Transaction Publishers, 1989.

HOMER-DIXON THOMAS F., *On the Threshold: Environmental Changes as Causes of Acute Conflict* in «International Security», 16, n. 2, autunno 1991.

I buchi neri e il paradosso della forza centrifuga

Per un oggetto che orbiti nelle vicinanze di un buco nero la forza centrifuga diviene in realtà centripeta, un effetto solo apparentemente paradossale che ha implicazioni di grande importanza per l'astrofisica

di Marek Artur Abramowicz

Chiunque si sia mai trovato a bordo di un'automobile, un autobus o un treno che imbocca una curva ad alta velocità ha sperimentato la forza centrifuga, quella spinta verso l'esterno della curva che diventa più intensa al crescere della velocità del veicolo. È facile quindi immaginare la sorpresa mia e del mio collega A. R. Prasanna, del Physical Research Laboratory di Ahmedabad in India, quando ci siamo recentemente resi conto che, secondo la teoria della relatività generale di Einstein, in alcune circostanze la forza centrifuga può essere centripeta: può cioè puntare verso il centro di un moto circolare anziché verso la periferia. Abbiamo dimostrato che se un astronauta riuscisse a condurre la propria astronave abbastanza vicina a un oggetto estremamente compatto di grande massa, come un buco nero, egli avvertirebbe una forza «centrifuga» diretta verso l'interno e non verso l'esterno; al contrario di quanto si verifica ordinariamente, l'aumento della velocità orbitale porterebbe a un aumento della forza centripeta.

Nella regione vicina a un buco nero, secondo i calcoli che abbiamo effettuato, non solo la forza centrifuga cambia verso, ma si capovolgono tutti gli effetti dinamici che configurano una distinzione tra interno ed esterno, centro e periferia. Questa scoperta è importante per capire alcuni aspetti della fisica dei buchi neri, che si pensa siano una componente essenziale di quel misterioso motore centrale che alimenta le galassie più brillanti del cosmo. Lo studio del paradosso della forza centrifuga ha fornito alcuni indizi estremamente interessanti sul comportamento di queste fonti di energia galattiche.

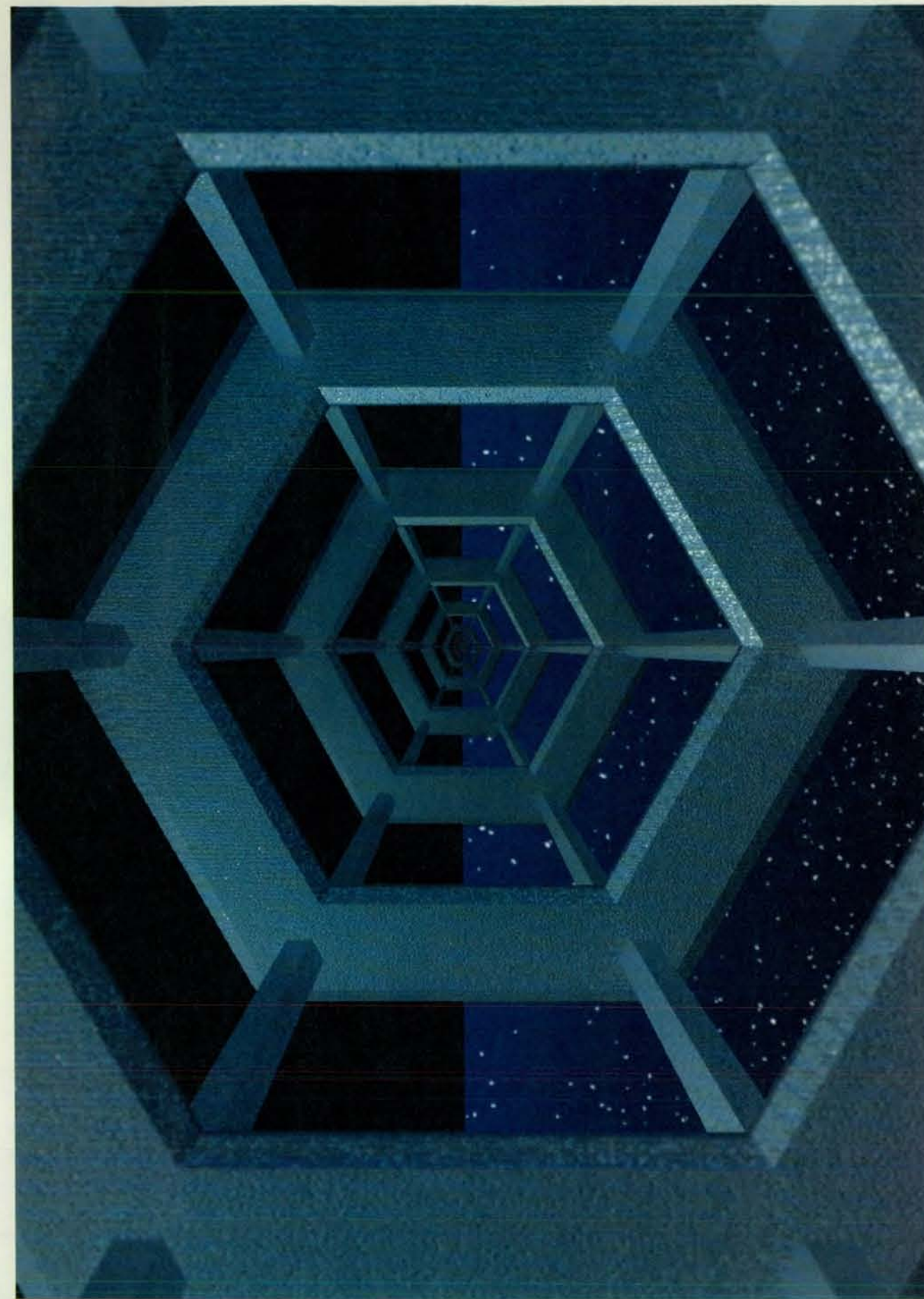
La causa del paradosso della forza centrifuga è l'incredibile intensità del campo gravitazionale prodotto dai buchi neri. Come aveva previsto nel 1915 Albert Einstein, un campo gravitazionale deforma lo spazio e deflette la luce. Nel 1919 Sir Arthur Stanley Eddington confermò questa previsione misurando la minuscola deviazione subita dalla luce di sorgenti lontane quando passa in prossimità del Sole, che è in grado di produrre una deflessione inferiore a un millesimo di grado per un fascio luminoso che sfiori la superficie. I buchi neri, che generano un campo gravitazionale di gran lunga più intenso di quello solare, piegano la luce in misura corrispondentemente maggiore.

Gli astronomi non hanno mai osservato direttamente un buco nero, ma si sono raccolte prove indiziarie sufficienti a convincere quasi tutta la comunità scientifica della loro esistenza. Nell'ultimo ventennio, infatti, sono stati identificati molti oggetti che sembrano candidati a contenere un buco nero: varie sorgenti brillanti di raggi X nella Galassia e molti nuclei galattici attivi, ossia nuclei eccezionalmente luminosi di galassie lontane.

Un buco nero imprigiona per sempre qualunque tipo di radiazione o materia vi passi troppo vicino. Il punto di non ritorno definisce l'estensione, o raggio gravitazionale, del buco nero. Un buco nero con una massa equivalente a quella del Sole avrebbe un raggio gravitazionale di circa tre chilometri. Se, per esempio, un raggio di luce si propagasse parallelamente alla superficie del buco nero a una distanza pari a tre volte il raggio gravitazionale, risulterebbe curvato di circa 45 gradi. Un fatto di particolare rilievo è che un raggio di luce che lambisca il buco nero a una distanza esattamente pari a 1,5 raggi gravitazionali vi girerà intorno in un'orbita perfettamente circolare. L'esistenza di questi raggi luminosi circolari è un elemento fondamentale del paradosso della forza centrifuga.

Quasi 20 anni fa Jean-Pierre Lasota (che oggi lavora all'Osservatorio di Parigi) e io scoprimmo per puro caso i primi indizi dell'esistenza del paradosso. Stavamo lavorando presso il Centro astronomico Copernico di Varsavia su un problema piuttosto tecnico di teoria della relatività generale. Per l'esattezza eravamo alle prese con una formula complessa ricavata da una nostra studentessa, Bożena Muchotrzeb. Doveva evidentemente esserci un errore: la formula calcolava la forza centrifuga subita da un oggetto orbitante intorno a un buco nero sulla stessa traiettoria dei raggi luminosi circolari, e stabiliva che l'oggetto, a qualsiasi velocità si muovesse, avrebbe sempre avvertito la stessa forza totale che lo attraeva verso il centro. In particolare, un oggetto immobile avrebbe sperimentato la stessa forza centripeta subita da un proiettile in moto a una velocità prossima a quella della luce.

Per noi non poteva trattarsi che di un'assurdità pura e semplice. Secondo la dinamica elementare la forza centrifuga dipende dalla velocità orbitale, ma la

Una struttura spaziale di travi e centine esagonali si stende intorno a un buco nero sferico a una quota pari a 1,5 volte il raggio del buco nero. La struttura s'incurva intorno al buco, ma a un osservatore posto al suo interno apparirebbe rettilinea. Il fenomeno è dovuto al fatto che, proprio a quella quota, il campo gravitazionale del buco nero ha una intensità tale da far percorrere ai raggi luminosi traiettorie perfettamente circolari. Un osservatore orbitante intorno al buco nero lungo questa struttura non avvertirebbe forza centrifuga. Anche la lieve distorsione delle centine lontane è dovuta alla curvatura della luce.

forza gravitazionale no. La forza totale - che altro non è se non la somma di quella gravitazionale e di quella centrifuga - deve quindi dipendere anch'essa dalla velocità orbitale. Dato che la formula non dava la risposta che ci attendevamo, eravamo fermamente convinti che non potesse proprio essere corretta. Eppure, ripetendo attentamente tutti i calcoli necessari per ricavarla, non riuscivamo a trovare errori. Alla fine risultò che la formula era giusta, e quindi doveva essere corretta anche la sua previsione paradossale sul comportamento della materia che percorra la traiettoria di un raggio di luce circolare.

Non ci sono veri paradossi in fisica. A volte si trova paradossale un fenomeno per inerzia mentale, quando si rimane ancorati a una concezione incompleta che impedisce di capire come vadano le cose nella realtà. Lasota e io ci rendemmo infine conto che il moto sulla traiettoria di un raggio di luce circolare sembra essere un paradosso così stridente perché è difficile accettare il fatto che tali raggi, pur essendo circolari, in un certo senso sono anche perfettamente rettilinei.

Per elaborare un'immagine intuitiva appropriata dei raggi di luce circolari si pensi a due astronauti (chiamiamoli Alice e Bob) che svolgono esperimenti in una stazione spaziale costruita intorno a un buco nero. La stazione ha la forma di un tubo circolare centrato esattamente sulla traiettoria del raggio di luce circolare, che quindi coincide con l'asse del tubo. Gli astronauti sanno che l'asse del tubo è circolare perché Bob ha misurato la curvatura delle pareti per tutta la sua lunghezza usando regoli diritti; eppure, a causa della deflessione dei raggi luminosi, essi vedono il tubo perfettamente rettilineo.

Per capire come questo avvenga, immaginiamo che Alice fissi un riflettore al centro della sezione del tubo e poi cominci a camminare lungo il tubo stesso, allontanandosi dal riflettore. La sorgente luminosa le apparirà sempre al centro della sezione del tubo, senza mai sparire dietro la curva; dovunque ella si trovi, con la sua traiettoria circolare la luce la raggiungerà sempre. Guardandosi alle spalle, Alice vedrà il riflettore sempre più fievole via via che si allontana, mentre davanti a sé lo vedrà sempre più luminoso; in realtà la luce circola lungo il tubo ripetutamente, e quindi Alice ne vedrà addirittura un'immagine multipla.

Alice potrebbe avere qualche difficoltà a spiegare perché vede il riflettore sia davanti sia dietro di sé, ed essere confusa alla vista dell'immagine multipla, ma deve comunque concludere che il tubo è rettilineo perché le pareti non nascondono mai il riflettore. Sulla base di ciò che vede, quindi, Alice non si aspetterebbe che sugli oggetti in moto lungo il tubo agiscano effetti centrifughi e dedurrebbe che su questi oggetti la forza centrifuga sia nulla. Supporrebbe inoltre che l'unica forza attiva sugli oggetti nel tubo sia quella gravitazionale, che non dipende dalla velocità orbitale. Ora, il fatto interessante è che Alice può formulare previsioni esatte proprio sulla base di ciò che vede: è quello che io chiamo principio del «vedere per credere».

Il significato profondo di questo principio non mi fu del tutto chiaro fino al 1985. Quell'anno, un giorno di primavera, tenni una chiacchierata informale all'ora di pranzo sul paradosso del raggio di luce circolare presso l'Institute for Theoretical Physics di Santa Barbara, in California. Avevo la fortuna di parlare al cospetto di vari esperti di relatività, tra i quali Brandon Carter dell'Osservatorio di Parigi. Il giorno successivo Carter escogitò un'idea brillante: si era reso conto che se un oggetto si muove a velocità costante sulla traiettoria di un raggio di luce qualsivoglia - rettilineo, curvo o circolare che sia - la forza che lo tiene «in pista» non dipende dalla velocità del moto. Naturalmente l'oggetto segue la traiettoria del raggio nello spazio, ma la sua velocità è inferiore a quella della luce.

Se per esempio un razzo dovesse seguire la traiettoria di un raggio di luce che rasenta il Sole, dovrebbe modificare progressivamente la propria direzione di un millesimo di grado. Per mantenere questa traiettoria conservando una velocità costante, dovrebbe azionare i retrorazzi in direzione perpendicolare alla traiettoria stessa; tuttavia la forza che questi dovrebbero esercitare non dipenderebbe dalla velocità del razzo.

Carter avanzò l'ipotesi che il principio del «vedere per credere» valga ovunque, in qualsiasi campo gravitazionale. In altre parole, un oggetto qualsiasi che si muova a velocità costante lungo la traiettoria di un raggio di luce curvato da un campo gravitazionale si comporta come se si muovesse in linea retta. Successivamente Carter, Lasota e io dimostrammo che questa ipotesi è giusta, a condizione che il campo gravitazionale in questione non si modifichi nel tempo. Elaborammo così il concetto di «geometria ottica», una struttura concettuale molto utile per studiare il comportamento dinamico di oggetti in campi gravitazionali intensi. In seguito John C. Miller dell'Osservatorio astronomico di Trieste e Zdĕnek Stuchlík dell'Università sleslana di Opava, in Moravia, scoprirono nell'ambito di questo quadro concettuale alcune relazioni fondamentali tra dinamica e geometria, e Norbert Wex del Max-Planck-Institut für Physik und Astrophysik di Monaco di Baviera propose un sistema ingegnoso ed elegante per adattare la geometria ottica al caso di buchi neri in rotazione.

La geometria spaziale convenzionale si basa su misurazioni effettuate con normali «regoli» rettilinei, che definiscono l'unità di lunghezza. La geometria ottica, invece, si basa su misurazioni effettuate tramite segnali luminosi.

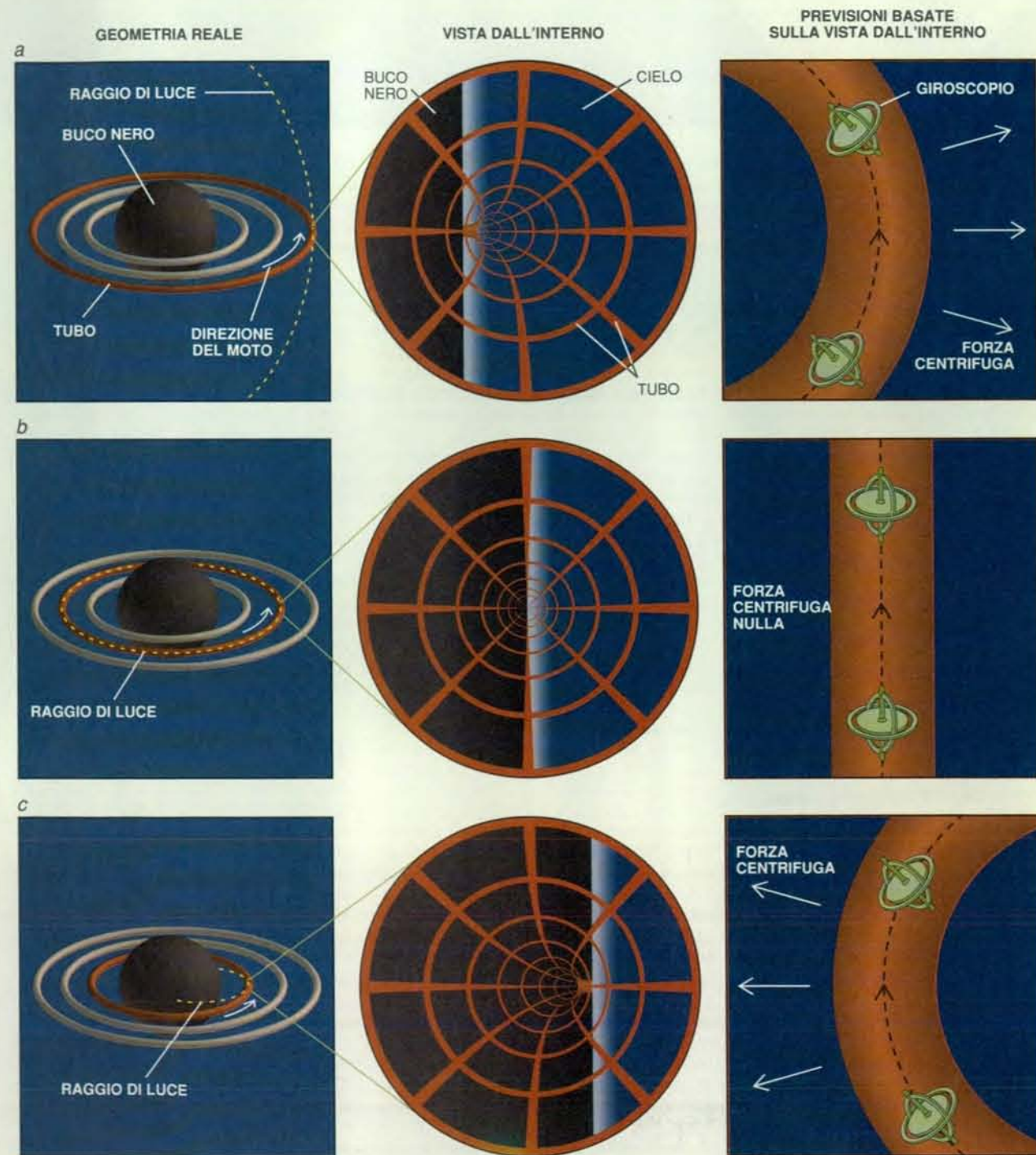
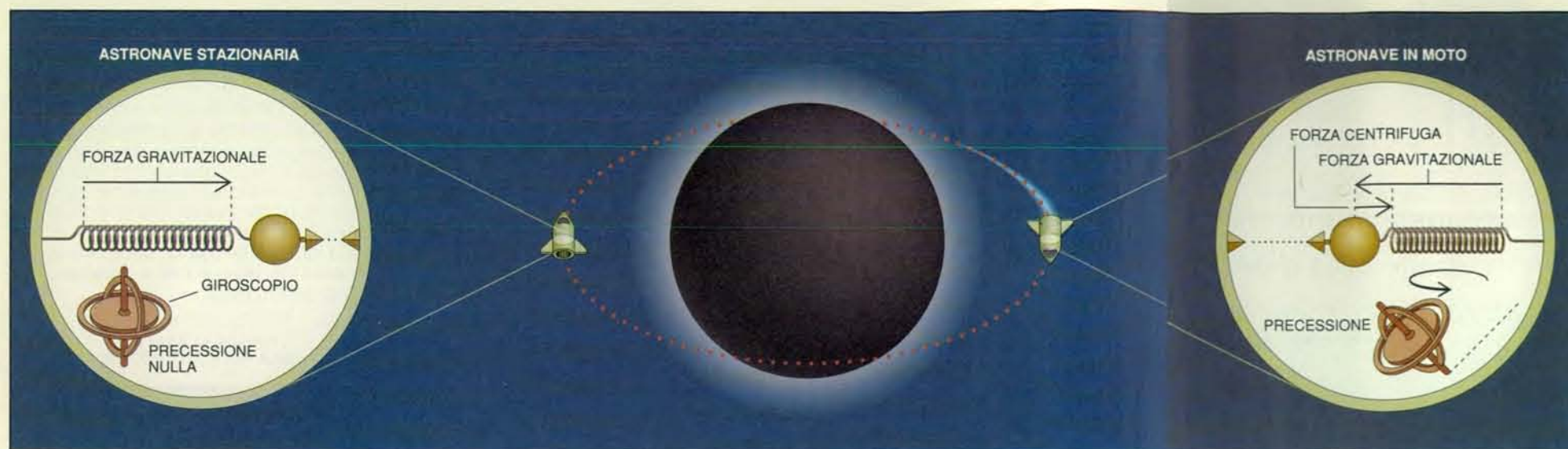
Nella geometria convenzionale si può misurare la lunghezza di una curva contando quanti regoli la ricoprono. Si può allora definire la distanza tra due punti nello spazio come la lunghezza della curva più breve che li unisce; questa curva di lunghezza minima si chiama geodetica. Se si fanno misurazioni in uno spazio piatto o, in alternativa, in uno spazio senza campi gravitazionali, la curva più breve tra due punti, la geodetica, è semplicemente una linea retta.

In geometria ottica invece la distanza tra due punti nello spazio è definita come la metà del tempo che la luce impiega per andare da un punto all'altro e tornare, misurato con un orologio situato nel primo punto. In uno spazio privo di campi gravitazionali la geometria ottica è identica a quella convenzionale perché sia i raggi luminosi sia le geodetiche sono linee rette. In questo caso quindi i raggi luminosi delineano la geometria dello spazio.

Secondo la teoria della relatività generale di Einstein, però, le tre dimensioni dello spazio e la singola dimensione del tempo formano uno spazio-tempo quadridimensionale. In qualsiasi spazio-tempo, in cui siano presenti o meno campi gravitazionali, la luce si muove sempre lungo geodetiche, e quindi delinea in ogni caso la geometria dello spazio-tempo. Se lo spazio è deformato da un campo gravitazionale, però, i raggi di luce sono curvi e in generale non seguono le sue geodetiche (ciò significa che le geodetiche di uno spazio non coincidono con quelle dello spazio-tempo di cui fa parte). Quindi, in generale, i raggi di luce non delineano la geometria del solo spazio.

La geometria ottica ripristina la corrispondenza tra geometria dello spazio e traiettorie dei raggi luminosi applicando una trasformazione a tutte le distanze «vere» (quelle misurate con regoli rettilinei). L'impiego della geometria ottica è simile, sotto molti aspetti, al procedimento che si segue per realizzare una carta piana di un globo sferico; si tratta cioè di un sistema per ottenere mappe maneggevoli di uno spazio curvo. Come tale, la geometria ottica va incontro ad alcune delle stesse difficoltà della cartografia convenzionale, in particolare l'impossibilità di rappresentare un globo su una carta piana senza introdurre una qualche distorsione. Tanto in cartografia quanto in geometria ottica una determinata rappresentazione può minimizzare l'aberrazione di certe caratteristiche, ma al prezzo di uno stravolgimento totale di altre. La scelta della rappresentazione dipende quindi dallo scopo per cui si realizza la carta. La ben nota proiezione geografica di Mercatore, per esempio, ingrandisce enormemente le regioni polari, ma è preziosissima per la navigazione perché rappresenta come

La forza centrifuga vicino a un buco nero

A un osservatore lontano, tre tubi costruiti intorno a un buco nero apparirebbero circolari, ma non è detto che un osservatore che si trovi al loro interno sia d'accordo. Il primo tubo è abbastanza lontano dal buco nero, in una zona in cui la luce viaggia quasi in linea retta. In questo caso tutti e due gli osservatori lo vedrebbero curvare intorno al buco nero, e tutti e due concluderebbero correttamente che un oggetto in moto lungo il tubo venga spinto verso l'esterno, in direzione opposta al buco nero, dalla forza centrifuga. Un giroscopio che percorra il tubo precederebbe sotto l'effetto della stessa forza. Il secondo tubo è costruito invece in quella particolare regione di spazio dove il campo gravitazionale del buco nero è tale da piegare i raggi luminosi in una traiettoria perfettamente circolare. Dato che la luce segue un percorso curvo lungo il tubo, un osservatore al suo interno lo considererebbe perfettamente diritto e riterrebbe, a ragione, che non ci sia alcuna forza centrifuga. Il terzo tubo è infine ancora più vicino al buco nero. In questo caso i raggi luminosi si piegano a tal punto che il tubo sembra curvare in direzione opposta al buco nero. Un osservatore nel tubo prevede correttamente che la forza centrifuga sia diretta verso l'interno, in direzione del buco nero, e che il giroscopio preceda sotto l'azione di questa forza.

Per misurare la forza centrifuga si possono usare due astronavi che seguono la stessa orbita intorno a un buco nero. Ciascuna è dotata di un giroscopio e di un peso attaccato a una molla. Ogni astronave manovra, senza deviare dall'orbita assegnata, in modo che il peso punti verso un certo segno di riferimento sullo scafo. Poi una delle due astronavi modifica la velocità orbitale fino a fermarsi, cioè finché il giroscopio non smette di precedere. A questo punto la forza centrifuga deve essere nulla, e quindi la forza totale che agisce sul peso deve essere pari a quella gravitazionale. L'altra astronave si muove a una velocità arbitraria. Si può ricavare la forza centrifuga cui è sottoposto il peso misurando la tensione della molla e confrontando il risultato con quello ottenuto dalla prima astronave.

rette tutte le lossodromiche (curve sul globo che incontrano con angolo costante i successivi paralleli). Analogamente, la geometria ottica distorce le distanze effettive, ma è utilissima per studiare la propagazione della luce e i fenomeni dinamici, perché nelle carte ottenute in questo modo i raggi luminosi sono geodetiche. (Naturalmente questo vale a condizione che il campo gravitazionale non vari nel tempo e che la sua sorgente non ruoti su se stessa.) In geometria ottica propagazione della luce e fenomeni dinamici sono collegati, contrariamente a quanto avviene nella geometria convenzionale dello spazio.

La trasformazione delle distanze usata in geometria ottica è un esempio di trasformazione conforme, un tipo di procedura matematica impiegata spesso nella teoria della relatività. In particolare questa trasformazione «raddrizza» i raggi di luce curvi in modo da farli diventare geodetiche.

Applicando la geometria ottica si possono tralasciare alcune complessità tecniche imposte dalla curvatura dello spazio per concentrarsi invece sui problemi fisici fondamentali; questo tipo di trasformazione conforme permette così di comprendere in maniera intuitiva la dinamica degli spazi curvi. I fenomeni dinamici infatti concordano sempre con ciò che si osserva. La geometria ottica spiega quindi perfettamente il comportamento in apparenza paradossale degli oggetti che si muovono lungo la traiettoria di un raggio di luce circolare.

Il risultato generale più importante ottenuto con l'ausilio della geometria ottica è forse la scoperta che in certe situazioni lo spazio risulta rivoltato come un guanto, con l'esterno e l'interno scambiati. Me ne resi conto leggendo un articolo piuttosto tecnico di Malcolm Anderson e José P. S. Lemos, che fanno parte del gruppo di Donald Lynden-Bell dell'Università di Cambridge, i quali hanno dimostrato che, se una nube di gas percorre un'orbita molto stretta intorno a un buco nero, l'attrito viscoso nella nube trasferisce momento angolare verso l'interno. Questa scoperta è molto strana perché di norma l'attrito viscoso trasferisce momento angolare verso l'esterno.

In effetti il trasferimento verso l'esterno del momento angolare per effetto dell'attrito viscoso è un meccanismo di importanza assolutamente fondamentale per l'astrofisica, perché contribuisce a spiegare come il disco di accrescimento (un tipo particolare di nube di gas) in orbita intorno a un buco nero gli fornisca l'energia che serve ad alimentare i nuclei attivi di alcune galassie. L'attrito tra regioni interne ed esterne tende a rendere più rigida la rotazione del disco di accrescimento, rallentando le regioni interne che ruotano rapidamente e accelerando quelle esterne, più lente. Quando questo avviene il momento angolare si trasferisce dall'interno all'esterno.

Anderson e Lemos scoprirono che l'attrito viscoso poteva trasportare momento angolare verso l'interno, ma senza fornire una spiegazione convincente. Dopo aver letto il loro articolo, mi resi improvvisamente conto del fatto che la geometria ottica propone una spiegazione di grande respiro di questo effetto e di molti altri risultati ugualmente sorprendenti. Scoprii cioè che lo spazio vicino a un buco nero è rivoltato come un guanto: la direzione interno-esterno definita dai regoli diritti è esattamente opposta a quella definita dai raggi luminosi. Nella situazione descritta da Anderson e Lemos il momento angolare in realtà viene trasportato verso l'esterno, come è giusto che sia, ma il concetto di esterno va inteso alla luce della geometria ottica. Nella situazione relativamente più normale che si ha a una certa distanza dal buco nero l'orientazione interno-esterno della geometria ottica coincide con quella della geometria convenzionale, ma vicino al buco nero le due direzioni sono opposte, e quindi il momento angolare è trasferito verso quello che normalmente sarebbe l'interno, con un comportamento apparentemente paradossale.

Per capire come mai ciò avvenga, si pensi di nuovo alla stazione spaziale fatta a ciambella intorno al buco nero, dove Alice e Bob svolgono esperimenti. In questo caso però la stazione non è costruita intorno al raggio di luce circolare, ma segue un cerchio più piccolo, sempre centrato sul buco nero. Bob misura le distanze vere con un normale regolo diritto, mentre Alice usa segnali luminosi per le sue misurazioni. Supponiamo inoltre, per fissare le idee, che nel tubo i due astronauti siano rivolti in modo da mantenere sempre il buco nero alla propria sinistra. Con un regolo diritto, Bob riscontra che il tubo è curvo verso sinistra, e in effetti le sue misurazioni corrispondono alla geometria reale: se toccasse la parete interna del tubo con le mani, la sentirebbe curvare verso sinistra; ne desume perciò che l'esterno stia a destra.

Dall'esperienza quotidiana Bob sa inoltre che la forza centrifuga è diretta verso l'esterno. Prevede quindi che gli oggetti in moto lungo il tubo vengano spinti verso destra e, analogamente, che l'attrito viscoso trasporti il momento angolare a destra. Ma in realtà è vero l'esatto contrario.

Alice compie una serie di misurazioni diverse, basate su ciò che vede effettivamente, e queste la portano alla conclusione esatta. Per prima cosa chiede a Bob di allontanarsi portando con sé un riflettore in modo che questo percorra l'asse del tubo. A questo punto, se i raggi luminosi non venissero deflessi dal campo gravitazionale del buco nero (cioè se fossero rettilinei) la sorgente di luce scomparirebbe dietro la parete sinistra del tubo e Alice concluderebbe che esso curva a sinistra. Se i raggi fossero circolari il riflettore rimarrebbe sempre in vista, e quindi il tubo apparirebbe diritto; ma esso è tanto vicino al buco nero che i raggi luminosi sono ancora più curvi di quelli circolari. Alice vede perciò il riflettore sparire sulla destra e ne deduce che il tubo curva a destra. Su queste basi prevede che la forza centrifuga sia diretta verso sinistra e che l'attrito viscoso trasferisca momento angolare verso sinistra. Le sue previsioni sono giuste, come assicura il principio del «vedere per credere». Si noti che, in termini di geometria convenzionale, nel tubo la forza centrifuga attira verso il centro del moto circolare.

Grazie alla geometria ottica, negli ultimi anni si è riusciti a risolvere anche vari problemi astrofisici che coinvolgono il comportamento di materia che ruota su se stessa in campi gravitazionali molto intensi. I due più importanti problemi di questo tipo sono il collasso gravitazionale di stelle in rotazione e la fusione di due stelle di neutroni, che sono corpi estremamente densi, in un solo oggetto. John Miller e io abbiamo dimostrato che la geometria ottica può essere estremamente utile per affrontare questi problemi. Abbiamo fornito una spiegazione semplice della strana evoluzione della forma di una stella rotante che si contrae.

Impiegando la teoria non relativistica, ci si aspetterebbe che una massa rotante di gas che si contragga conservando il proprio momento angolare si appiattisca sempre più. Nel 1974 invece Subrahmanyan Chandrasekhar, dell'Università di Chicago, e Miller, che allora lavorava all'Università di Oxford, scoprirono che, secondo la teoria di Einstein, nelle ultime fasi della contrazione, quando il campo gravitazionale è molto intenso, l'appiattimento cessa e la stella diventa più tondeggiante. Miller e io scoprimmo la spiegazione corretta di questo effetto usando la geometria ottica e considerando il comportamento insolito della forza centrifuga in un campo gravitazionale così intenso.

Per convincere i miei colleghi che l'inversione della forza centrifuga è un fenomeno fisico reale c'è voluto del bello e del buono. Il nodo della discussione era se fosse possibile definire e misurare la forza centrifuga in uno spazio fortemente curvo. Questa domanda solleva diversi problemi molto raffinati che i miei critici e io abbiamo sviscerato svolgendo numerosi calcoli complessi. Gran parte dei progressi che ho compiuto derivano dagli sforzi fatti per rispondere alle obiezioni estremamente impegnative di Fernando de Felice dell'Università di Torino. In seguito alle mie amichevoli battaglie con de Felice, ho adottato una definizione particolare della forza centrifuga, che non è l'unica possibile, ma a mio avviso la più utile e convincente.

Ai fini della misurazione della forza centrifuga, immagino due astronavi che percorrano la stessa orbita intorno a un buco nero, pilotate rispettivamente da Alice e Bob. Ciascuna astronave ha a bordo due strumenti, un giroscopio e un peso fissato a una molla. Misurando la lunghezza della molla, gli astronauti possono determinarne la tensione, che a sua volta è pari alla somma delle due forze che agiscono sul peso: quella gravitazionale e quella centrifuga.

Per misurare una sola delle due forze, Alice e Bob devono cambiare l'orientazione della loro astronave rispetto all'orbita in maniera opportuna, facendola ruotare in modo che la molla estesa punti sempre verso un certo segno di riferimento sullo scafo. La direzione della molla è quindi fissa rispetto all'astronave, ma non rispetto allo spazio. Ognuno dei due giroscopi, al contrario, punta sempre in una direzione fissa nello spazio, e quindi presenterà un moto di precessione rispetto alla direzione della molla via via che l'astronave percorre la propria orbita.

Per misurare la forza gravitazionale Bob fa fermare la nave; sa quando questo avviene perché il giroscopio smette di precedere. A questo punto può affermare che la forza che estende la molla è puramente gravitazionale. Riferisce allora il suo risultato ad Alice, che continua a sfrecciare intorno al buco nero sulla stessa orbita. Alice misura la forza totale che estende la sua molla e ne ricava la forza centrifuga sottraendo la forza gravitazionale misurata da Bob. Questo metodo per misurare la forza centrifuga potrà sembrare complicato, ma ha il vantaggio di essere esattamente identico sia in campi gravitazionali forti sia in campi gravitazionali deboli.

Il valore pratico della geometria ottica sta nel fatto che fornisce un metodo comodo per trattare problemi difficili di relatività generale. Inoltre ha un'utilità didattica perché permette di capire intuitivamente vari effetti relativistici importanti per l'astrofisica moderna, che grazie a essa non appaiono più paradossali o incomprensibili.

A livello fondamentale, la geometria ottica mostra che i concetti di «dentro» e «fuori» non sono assoluti ma, in uno spazio deformato da intensi campi gravitazionali, relativi. Attualmente non ci sono difficoltà a ritenere che sinistra e destra, come pure alto e basso, siano concetti relativi, ma questi asserti non sono stati sempre così ovvi: in epoca biblica alcuni ritenevano che «destra» e «sinistra» fossero entità assolute, sulla base di alcuni passi dell'Antico Testamento e di altri testi antichi. Diversi secoli or sono si credeva che fossero concetti assoluti quelli di «alto» e «basso»: era difficile immaginare che gli abitanti degli antipodi camminassero capovolti. Forse, prima della fine del prossimo secolo, nessuno si sorprenderà che interno ed esterno siano relativi.

BIBLIOGRAFIA

ABRAMOWICZ MAREK ARTUR e LASOTA JEAN-PIERRE, *On Travelling Round without Feeling It and Uncurving Curves* in «American Journal of Physics», 54, pp. 936-939, ottobre 1986.

ALLEN BRUCE, *Reversing Centrifugal Forces* in «Nature», 347, n. 6294, 18 ottobre 1990.

ABRAMOWICZ MAREK ARTUR, *Relativity of Inwards and Outwards: An Example* in «Monthly Notices of the Royal Astronomical Society», 256, n. 4, 15 giugno 1992.

ABRAMOWICZ MAREK ARTUR e SZUSZKIEWICZ E., *The Wall of Death* in «American Journal of Physics» (in stampa).

La resistenza nei superconduttori ad alta temperatura

Questi materiali perdono le loro proprietà quando vengono sottoposti a intensi campi magnetici, un fenomeno la cui comprensione dovrebbe condurre verso applicazioni tecnologiche di grande interesse economico

di David J. Bishop, Peter L. Gammel e David A. Huse

Allorché nel 1986 vennero scoperti i superconduttori ad alta temperatura, la fantasia dei fisici si sbizzarrì in prospettive ambiziose, come la possibilità di progettare treni a levitazione magnetica o dispositivi ad alta capacità per l'immagazzinamento di energia elettrica. La scoperta in certi materiali ceramici della proprietà di condurre corrente elettrica con resistenza pressoché nulla anche ad alta temperatura aveva suscitato un entusiasmo che dovette però ben presto scontrarsi con diverse difficoltà.

Uno degli ostacoli più spiacevoli consiste nel fenomeno di distruzione dello stato di superconduzione che si verifica quando il materiale viene sottoposto a un campo magnetico. La presenza del campo magnetico è infatti una condizione irrinunciabile per la maggior parte delle applicazioni che erano state inizialmente previste. La resistenza al flusso di corrente può verificarsi quando il campo magnetico esterno penetra nel superconduttore in forma di tubi di flusso discreti, detti anche flussoidi. Dato che un flussoide consiste in un mulinello di corrente elettrica, esso viene spesso chiamato vortice. Se questi vortici si muovono, possono ostacolare il libero flusso degli elettroni; conoscerne pertanto con precisione la dinamica e il modo in cui si dispongono nelle diverse condizioni di temperatura e di campo magnetico è una condizione di importanza cruciale per il controllo di questo fenomeno e per il mantenimento nel materiale dello stato di superconduzione.

Fortunatamente alcuni recenti studi hanno di molto accresciuto le nostre conoscenze sui vortici. Si è scoperto che il comportamento dei superconduttori esposti a un campo magnetico è molto più complesso di quanto non si pensasse in passato. È stato infatti appurato che, nella famiglia dei superconduttori ad alta temperatura, i vortici sono in grado di dar luogo a tutta una serie di nuove fasi della materia. Per descrivere queste fasi - solidi, liquidi, vetri di vortici - si sono dovute abbandonare alcune ipotesi ritenute in precedenza valide nel campo della superconduttività e avanzarne altre basate su moderni concetti di fisica degli stati condensati. Per valutare queste nuove teorie sono state messe a punto tecniche sperimentali dalla sensibilità senza precedenti, e si spera che esse potranno indicare la via verso la completa comprensione e forse anche l'impiego pratico di questi nuovi materiali.

Ritornando al passato, non ci si deve stupire del fatto che le conoscenze sullo stato di superconduzione accumulate prima del 1986 fossero inadeguate alla descrizione della superconduttività ad alta temperatura. Le prime teorie si svilupparono partendo da osservazioni su superconduttori convenzionali.

Questi materiali, generalmente metalli e leghe abbastanza comuni, conducono elettricità senza resistenza solo quando vengono raffreddati a temperature di qualche grado sopra lo zero assoluto. In effetti fu la curiosità circa il comportamento della materia a basse temperature a condurre il fisico olandese Heike Kamerlingh Onnes alla scoperta della superconduttività nel 1911. Onnes giunse alla sua scoperta grazie al successo conseguito nel formidabile compito di liquefare l'elio, ultimo fra i gas inerti a essere condensato. L'elio liquido permise a Onnes di raffreddare materiali a temperature di poco più di un kelvin, vale a dire un grado centigrado sopra lo zero assoluto. (Lo zero assoluto è pari a −273,16 gradi Celsius.)

Secondo una storia forse non attendibile, la scoperta della superconduttività si verificò quando Onnes chiese a uno studente di misurare la resistenza elettrica del mercurio; nella sua relazione questi riferì che la resistenza scompariva quando la temperatura del campione scendeva a 4,2 kelvin. Onnes rispedì lo studente in laboratorio affinché individuasse quello che secondo lui doveva essere un «errore» che provocava un artefatto sperimentale. Dopo diverse prove, sempre con lo stesso risultato, i ricercatori si resero conto d'aver fatto una scoperta storica. Per questa e molte altre scoperte nel campo della fisica delle basse temperature, nel 1913 Onnes fu insignito del premio Nobel.

La resistenza nulla al flusso di corrente elettrica non era il solo motivo di meraviglia; il comportamento dei superconduttori in un campo magnetico si rivelò altrettanto sbalorditivo. Nel 1933 due fisici tedeschi, Karl Walther Meissner e Robert Ochsenfeld, scoprirono che un superconduttore può espellere il flusso magnetico che contiene quando viene raffreddato al di sotto della sua temperatura di transizione; la completa espulsione del flusso è nota oggi sotto il nome di effetto Meissner. Assieme all'assenza di resistenza, è stata proprio questa capacità di espellere il flusso magnetico a suscitare gli enormi interessi di ricerca nel campo della superconduttività.

A questo punto le osservazioni sperimentali avevano superato la teoria. I modelli quantistici sviluppati negli anni trenta potevano spiegare la conduttività nei metalli normali, ma non lo stato di superconduzione. Il problema si rivelò particolarmente ostico e i ricercatori non riuscirono a escogitare alcuna spiegazione teorica significativa sulle origini microscopiche della superconduttività fino agli anni cinquanta. Fu allora che due studiosi russi, Vitalj L. Ginzburg e Lev D. Landau, svilupparono una teoria fenomenologica. Osservando ciò che avveniva durante la transizione dallo stato normale a quello di superconduzione, i due scienziati riuscirono a formulare una serie di equazioni che descrivevano il fenomeno; non si riusciva però a spiegare perché questo si verificasse.

Nel 1957 John Bardeen, Leon N. Cooper e J. Robert Schrieffer misero a punto una teoria che fornì una spiegazione microscopica della superconduttività. Secondo la cosiddetta teoria BCS, gli elettroni di conduzione si propagano senza incontrare resistenza perché si muovono in coppie, le cosiddette coppie di Cooper. Gli elettroni formano coppie di Cooper in quanto interagiscono con i fononi, vibrazioni meccaniche del reticolo cristallino del metallo che assomigliano a onde acustiche. Le vibrazioni degli atomi del reticolo tendono a neutralizzare la repulsione che normalmente si esercita tra gli elettroni, anzi producono addirittura una debole forza attrattiva. L'intensità di questa interazione dipende fortemente dalla temperatura. Il punto, su una scala delle temperature, al quale si manifesta la superconduttività viene definito temperatura critica o tempe-

I vortici, qui rappresentati dai tubi verdi e rossi, sono «mazzi» discreti di linee di forza che penetrano in un superconduttore quando viene applicato un campo magnetico. L'immagine al calcolatore rappresenta l'intensità del campo magnetico (indicata come altezza dei tubi) sulla superficie del campione; il campo è massimo al centro di ogni vortice. La proiezione in basso mostra come i vortici (*punti bianchi*) si dispongano secondo un reticolo triangolare regolare all'interno del superconduttore.

ratura di transizione. A temperature al di sopra di questo punto critico, le fluttuazioni termiche distruggono le coppie di Cooper e di conseguenza lo stato di superconduzione del metallo.

L'interazione di accoppiamento determina due importanti scale microscopiche di distanza in un superconduttore. La prima di queste è la separazione spaziale degli elettroni di una coppia di Cooper: questa distanza è chiamata lunghezza di coerenza ed è la più piccola distanza in un superconduttore lungo la quale le proprietà elettroniche, come per esempio la resistività locale, possono cambiare. In superconduttori tipici, la lunghezza di coerenza può andare da alcune decine ad alcune centinaia di nanometri. (Queste scale di distanza sono correlate alla realtà atomica e possono essere di difficile comprensione a livello intuitivo; un nanometro equivale a 10^{-9} metri e gli atomi della maggior parte dei materiali distano fra loro 0,1-0,3 nanometri.)

La seconda lunghezza microscopica caratteristica è relativa alla forza dell'effetto Meissner, cioè alla capacità del superconduttore di espellere il flusso di un campo magnetico applicato. L'effetto si ha quando un piccolo campo magnetico viene applicato a un superconduttore, generando correnti che fluiscono presso la superficie del materiale. Queste correnti indotte creano un campo magnetico che cancella esattamente, nel resto del materiale, il campo applicato, e la loro intensità decade esponenzialmente all'aumentare della distanza dalla superficie del superconduttore. La distanza lungo la quale si verifica questo decadimento viene detta profondità di penetrazione ed è la distanza più breve lungo la quale un campo magnetico può variare in un superconduttore. In superconduttori tipici essa può andare da qualche decina fino a qualche migliaio di nanometri.

Queste lunghezze microscopiche definiscono due categorie molto diverse di superconduttori. Nei superconduttori di tipo I la lunghezza di coerenza è maggiore della profondità di penetrazione. Questi materiali tendono a essere superconduttori in condizioni di bassa temperatura e debole campo magnetico; se il campo raggiunge l'intensità critica (la quale varia da sostanza a sostanza), entra nel materiale distruggendo lo stato di superconduzione. Dato che nel caso dei superconduttori di tipo I la superconduzione scompare in presenza di campi magnetici di intensità relativamente bassa, questi materiali presentano uno scarso interesse dal punto di vista tecnologico.

I superconduttori di tipo II sono molto più utili. La profondità di penetrazione in questi materiali è maggiore della lunghezza di coerenza e pertanto essi rimangono nello stato di superconduzione anche dopo la penetrazione del campo magnetico. I superconduttori di tipo II possono sopportare campi molto forti - fino a quello che viene chiamato campo critico superiore - e quindi trasportare correnti più intense. Tutti i superconduttori di interesse tecnologico, inclusi quelli ad alta temperatura finora noti, sono di questo tipo.

Negli anni cinquanta il fisico russo Aleksej A. Abrikosov pubblicò la teoria di base sul comportamento di un superconduttore convenzionale di tipo II in presenza di un campo magnetico. Prendendo spunto dal lavoro di Ginzburg e Landau, Abrikosov riuscì a dimostrare che il comportamento magnetico di un superconduttore di tipo II al di sotto della temperatura critica dipende dall'intensità del campo applicato e dalla temperatura. Tale relazione può essere rappresentata da un diagramma di fase magnetico, il quale mostra che un superconduttore convenzionale ha tre stati magnetici distinti.

Il primo è lo stato Meissner, cioè lo stato in cui il materiale espelle totalmente il flusso magnetico applicato. Il superconduttore esiste in questo stato finché il campo magnetico resta al di sotto di una certa intensità: questo campo, chiamato campo critico inferiore, dipende in generale dalla temperatura.

Il secondo stato si presenta se il campo applicato viene portato a un livello più alto del campo critico inferiore. A questo punto il campo magnetico può ancora penetrare nel superconduttore, ma non completamente o uniformemente. La penetrazione nel materiale avviene per mezzo di tubi di flusso discreti che formano intrusioni tubolari del campo applicato. La meccanica quantistica dei superconduttori richiede che ogni tubo di flusso porti la stessa quantità di flusso magnetico: questa unità di flusso è nota come quanto di flusso. Dato che ogni flussoide deve portare la stessa quantità di flusso, ogni cambiamento del campo magnetico applicato deve modificare la densità dei flussoidi. In altre parole, al variare del campo applicato varierà la distanza fra i flussoidi. La configurazione di minima energia per una tale schiera di linee di flusso (vista stando al di sopra della superficie) è quella di un reticolo triangolare.

La struttura di una singola linea di vortice dipende dalla lunghezza di coerenza e dalla profondità di penetrazione. Ogni linea di flusso presenta un piccolo «nocciolo» centrale, il cui diametro è legato alla lunghezza di coerenza. All'interno del nocciolo il materiale è un metallo normale, mentre attorno a esso circolano le supercorrenti. (La corrente circolante è la ragione per la quale i fisici chiamano il nocciolo linea di vortice.) Queste supercorrenti producono un campo magnetico e la distanza lungo la quale il campo persiste è la profondità di penetrazione. È possibile visualizzare questi vortici usando minuscole particelle magnetiche; nelle immagini così ottenute risulta evidente l'ordinamento secondo un reticolo triangolare.

Al terzo e ultimo stato di un superconduttore si perviene se il campo magnetico applicato raggiunge un secondo

Il flusso di corrente attraverso un superconduttore può venire perturbato da vortici, ognuno dei quali consiste in un anello di corrente indotta da un campo magnetico. Le correnti applicate si sommano alle correnti circolanti da un lato del vortice, ma si sottraggono dall'altro. Ne risulta una forza che spinge i vortici perpendicolarmente al flusso di corrente; il movimento dissipa energia e produce resistenza.

La visualizzazione del reticolo dei flussoidi

Da bambini abbiamo tutti evidenziato le linee di forza di un magnete permanente usando un pezzo di carta e limatura di ferro. Alcuni di noi continuano a farlo: specificamente, è possibile visualizzare il campo magnetico che può permeare un superconduttore. Un piccolo campo magnetico entra nel superconduttore in tubi di flusso discreti, detti flussoidi o linee di vortice, che si ordinano in un reticolo regolare. Molte tecniche, fra cui la diffusione di neutroni e la microscopia a scansione a effetto tunnel, possono rivelare la configurazione del reticolo, ma probabilmente la visualizzazione magnetica usando particelle di ferro è ancora il metodo più semplice e diretto.

L'apparecchiatura impiegata (*a*), che misura 10 centimetri circa di altezza e 3 centimetri di diametro, consiste in pochi elementi chiave. Il superconduttore da studiare è collocato all'interno di una camera a vuoto riempita di elio. Applichiamo un campo magnetico dando energia alle bobine e raffreddiamo il campione al di sotto della sua temperatura di transizione. Riscaldiamo poi il filamento di tungsteno, al quale è unita una sferetta di ferro. La sferetta evapora e l'elio presente nella camera raffredda le particelle di ferro, producendo un «fumo» magnetico che si diffonde lentamente. Le particelle di ferro del fumo sono minuscole, circa 10 nanometri di diametro, e nel loro moto di deriva aggirano il diaframma che protegge il campione dal calore e vanno a coprire la superficie del superconduttore. Qui esse evidenziano le regioni dove le linee di vortice attraversano il materiale. Le particelle di ferro si «appiccicano» alla superficie, grazie alla debole forza di attrazione dovuta all'interazione di van der Waals, che agisce come una specie di «colla atomica». Il campione può poi essere riscaldato fino a temperatura ambiente e trattenere ancora le particelle di ferro; possiamo allora ottenere un'immagine diretta del reticolo utilizzando il microscopio elettronico.

Le linee di flusso appaiono come punti che rivelano la natura ordinata del reticolo (una fotografia del reticolo delle linee di vortice viene mostrata a pagina 36). Da queste immagini è possibile determinare la quantità di flusso magnetico per ogni linea di vortice. Questa grandezza è una costante fondamentale per i superconduttori, ed è nota come quanto di flusso, ϕ_0. Per tutti i superconduttori noti, $\phi_0 = hc/2e$ dove h è la costante di Planck, c è la velocità della luce ed e è la carica dell'elettrone. Il 2 a denominatore rappresenta una diretta conseguenza del fatto che gli elettroni in un superconduttore si muovono a coppie. Agli albori della ricerca nel campo dei superconduttori ad alta temperatura si pensava che il quanto di flusso potesse avere un valore diverso per questi materiali. Esperimenti come quelli che stiamo descrivendo, con i quali si può facilmente contare il numero di linee di vortice, esclusero rapidamente tale possibilità. Con un semplice conteggio si dimostra che il rapporto fra campo magnetico applicato e densità di linee di vortice è uguale al quanto di flusso.

Gli esperimenti di visualizzazione ci hanno permesso di individuare molte altre strutture. L'ordinamento delle linee di vortice è differente se il campo applicato incide sul campione con un angolo diverso da zero rispetto a uno degli assi cristallografici principali: invece di un reticolo regolare si ha la comparsa di catene di flusso (*b*).

Queste immagini consentono diversi tipi di analisi quantitative. Dopo aver digitalizzato la posizione delle linee di flusso, un calcolatore può unire con rette tutti i punti del reticolo (*c*). In un reticolo triangolare perfetto ogni linea di vortice possiede sei primi vicini e i difetti del reticolo appaiono come linee di vortice con un numero diverso di primi vicini; nell'immagine i difetti sono stati colorati in rosso per rendere l'identificazione più agevole.

Queste immagini di aspetto decorativo mostrano che il reticolo di flussoidi può assumere un particolare tipo di ordinamento, il cosiddetto ordinamento esagonale. Nell'ambito di strutture ordinate di questo tipo, la posizione delle particelle può essere casuale, ma gli angoli di legame fra primi vicini sono simili. Nell'ordinamento qui illustrato gli angoli di legame sono più o meno gli stessi da un'estremità all'altra della figura ma, a causa dei difetti, le particelle sono spaziate equamente solo per piccole distanze. Il modo più semplice per osservare questo ordinamento degli angoli di legame consiste nell'avvicinare il bordo della pagina all'occhio e traguardare le file di punti.

I diagrammi di fase magnetici mostrano che cosa accade quando un superconduttore di tipo II viene immerso in un campo magnetico. I superconduttori convenzionali presentano tre fasi (*a*). Nel cosiddetto stato Meissner si ha la completa espulsione del flusso magnetico applicato. Nello stato misto (o solido di vortici) il campo penetra nel superconduttore con tubi di flusso discreti, altrimenti detti flussoidi o linee di vortice. Nello stato normale la superconduzione viene meno e il campo penetra uniformemente nel materiale. I superconduttori ad alta temperatura presentano fasi simili (*b*), fatta eccezione per un regime a liquido di vortici. Questo stato esiste perché le fluttuazioni termiche «fondono» il solido di vortici, che può essere o un reticolo regolare (quando il superconduttore è pulito) o un vetro (se il superconduttore è sporco).

e più alto punto critico. Al di sopra di questo valore (campo critico superiore) lo stato di superconduzione viene completamente distrutto e il materiale è ricondotto al suo stato normale. La distruzione si verifica perché l'incremento dell'intensità del campo magnetico forza i vortici gli uni sempre più vicini agli altri. Quando i noccioli dei vortici, che si comportano come metalli normali, si sovrappongono troppo, non c'è più spazio a sufficienza fra di essi per mantenere la superconduttività.

La descrizione dei tre stati magnetici sembrava spiegare in modo sufficientemente dettagliato gli effetti di campi magnetici applicati a superconduttori. Nel 1986 però J. Georg Bednorz e K. Alex Müller del Centro di ricerche dell'IBM a Zurigo trovarono una nuova classe di superconduttori di tipo II. Si trovò che questi materiali, una famiglia di ceramiche a base di ossido di rame, potevano in certi casi essere superconduttori a temperature superiori a 120 kelvin, mentre per i superconduttori convenzionali le temperature critiche vanno da 20 a 25 kelvin. I superconduttori ad alta temperatura galvanizzarono il mondo scientifico: questi materiali potevano facilmente essere raffreddati con azoto liquido, che all'ingrosso costa circa 50 volte di meno dell'elio liquido usato per raffreddare i superconduttori convenzionali. Perfino piccoli refrigeratori di laboratorio consentono di scendere al di sotto della temperatura di transizione dei nuovi superconduttori.

Per quanto le elevate temperature critiche di questi materiali fossero estremamente interessanti, un fatto poco incoraggiante venne alla luce quando si studiarono le loro proprietà in funzione del campo magnetico applicato: i superconduttori ad alta temperatura non si comportavano secondo il modello di Abrikosov. Le discrepanze furono scoperte quando si studiarono i materiali in questione in presenza dei campi magnetici che sarebbero necessari per un impiego tecnologico, campi che possono raggiungere un'intensità anche di 10 tesla (un tesla è pari a circa 20 000 volte l'intensità del campo magnetico medio terrestre). In campi di questa intensità la resistenza di alcuni superconduttori non scendeva al di sotto di quella del rame fino a quando la temperatura non diminuiva fino al 20-30 per cento della temperatura di transizione. In certi casi la resistenza di alcuni materiali esposti a un campo magnetico rimaneva 100 volte più alta di quella del rame. I vantaggi della superconduttività ad alta temperatura sembravano persi, ma ulteriori esperimenti scoprirono la ragione di questo comportamento. Le linee di vortice si comportavano in modo strano, cioè non si ordinavano sempre in un reticolo triangolare rigido. Al contrario si scoprì che il reticolo di vortici si poteva «fondere» in un altro stato, un liquido di vortici; questo comportamento sopprimeva la transizione del materiale alla superconduttività.

Ci sono molte ragioni per le quali questo nuovo stato della materia dovrebbe ostacolare il flusso di corrente nei superconduttori ad alta temperatura. Forse il modo più semplice per capire questo effetto è di immaginare le linee di vortice in un superconduttore come bande elastiche. Una linea di vortice, proprio come un comune elastico, tende a mantenersi corta: allungarla impone infatti un certo dispendio di energia. Le fluttuazioni termiche, tuttavia, si oppongono a tale tendenza, in quanto fanno sì che sia gli atomi di un solido sia le linee di vortice vibrino con ampiezze crescenti al crescere della temperatura. Le linee di vortice, dunque, si allungano, mentre l'energia di una linea di vortice tende a riportarla nel suo stato non allungato.

Questa forza «di richiamo» è funzione della lunghezza di coerenza e della profondità di penetrazione. Una grande lunghezza di coerenza o una piccola profondità di penetrazione producono una forza elevata e quindi limitano le vibrazioni termiche delle linee di vortice. La maggior parte dei superconduttori ordinari di tipo II possiede questa caratteristica. La forza di richiamo è predomi-

nante e quindi mantiene le linee di vortice corte e diritte. Ne deriva che le fluttuazioni termiche delle linee di vortice sono piccole.

D'altra parte, i superconduttori ad alta temperatura possiedono caratteristiche pressoché opposte, cioè lunghezze di coerenza piccole e profondità di penetrazione grandi. La lunghezza di coerenza può essere anche minore di un nanometro, cioè da 10 a 100 volte inferiore a quelle dei superconduttori ordinari. La profondità di penetrazione dei superconduttori ad alta temperatura va da 100 a più di 10 000 nanometri, valori che superano quelli dei superconduttori ordinari di un fattore compreso fra 10 e 100.

La piccola lunghezza di coerenza e la grande profondità di penetrazione, insieme con le alte temperature di transizione, determinano, nei superconduttori ad alta temperatura, grandi fluttuazioni termiche delle linee di vortice. Invero, a temperature sufficientemente alte le linee vibrano abbastanza da «fondere» il reticolo di vortici. Il fenomeno è simile a quello della fusione del ghiaccio in acqua, della quale è responsabile l'agitazione termica delle molecole d'acqua. Per alcuni superconduttori ad alta temperatura, lo stato di liquido di vortici si mantiene in un intervallo di temperature più ampio di quello in cui si ha il reticolo di vortici.

Perché il liquido di vortici modifica la resistenza di un superconduttore? La risposta si trova pensando a quanto avviene quando si fa scorrere una corrente attraverso un superconduttore di tipo II immerso in un campo magnetico. Si ricordi che ogni linea di vortice consiste in un flusso di correnti circolanti attorno a un nocciolo centrale che si trova nello stato normale (non di superconduzione). Quando la corrente applicata fluisce nel campione, essa si somma alle correnti circolanti da un lato del vortice, mentre si sottrae dall'altro. Conseguentemente, sulla linea di vortice si esercita una forza che tende a far muovere il vortice stesso in direzione perpendicolare sia alla linea di vortice sia alla corrente applicata: questa forza si chiama forza di Magnus. Essa è simile alla portanza generata dall'ala di un aereo, una situazione nella quale l'aria fluisce più rapidamente lungo la faccia superiore del profilo alare che non lungo la faccia inferiore. Le linee di vortice sollecitate dalla forza di Magnus si muovono e così facendo dissipano l'energia del flusso di corrente; specificamente la dissipazione induce una differenza di potenziale e quindi una resistenza nel campione.

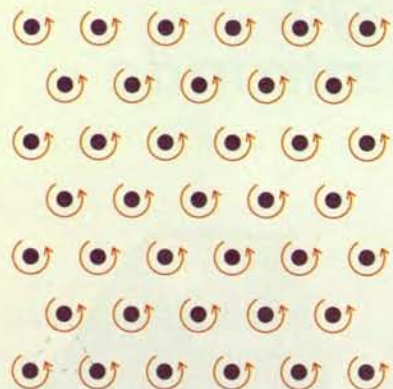
La misurazione di questa resistenza dimostra che il liquido di vortici si comporta come l'acqua in prossimità del punto di fusione. Abbiamo studiato la resistenza di un campione molto pulito di un superconduttore ad alta temperatura, $YBa_2Cu_3O_7$ (spesso abbreviato in YBCO), in funzione della temperatura, in presenza di un campo magnetico costante. Ad alta temperatura (cioè nella fase di liquido di vortici) la resistenza trovata è alta. Abbassando la temperatura, si è congelato il liquido di vortici e si è verificata la transizione allo stato di reticolo di vortici; in questo stato le linee di vortice non erano più libere di muoversi, cosicché la resistenza scompariva.

Le misurazioni di resistenza mostrano anche che il liquido di vortici viene lievemente sottoraffreddato prima di congelare; ciò è analogo a quanto si riscontra nel caso dell'acqua distillata, dove la fase liquida può persistere un poco al di sotto del punto di congelamento. Il sottoraffreddamento si può esprimere in termini più tecnici come segue: il comportamento della sostanza durante il riscaldamento non ricalca perfettamente quello riscontrato durante il raffreddamento; processi fisici di questo tipo sono detti di isteresi.

Tuttavia la descrizione generale del comportamento del liquido di vortici e della transizione a reticolo lascia ancora aperta una questione di fondamentale importanza ai fini delle applicazioni. Il liquido di vortici congela in un reticolo regolare solo se il materiale è pulito. Che cosa succede se il materiale è «sporco», cioè se nel reticolo atomico sono presenti impurezze chimiche o difetti?

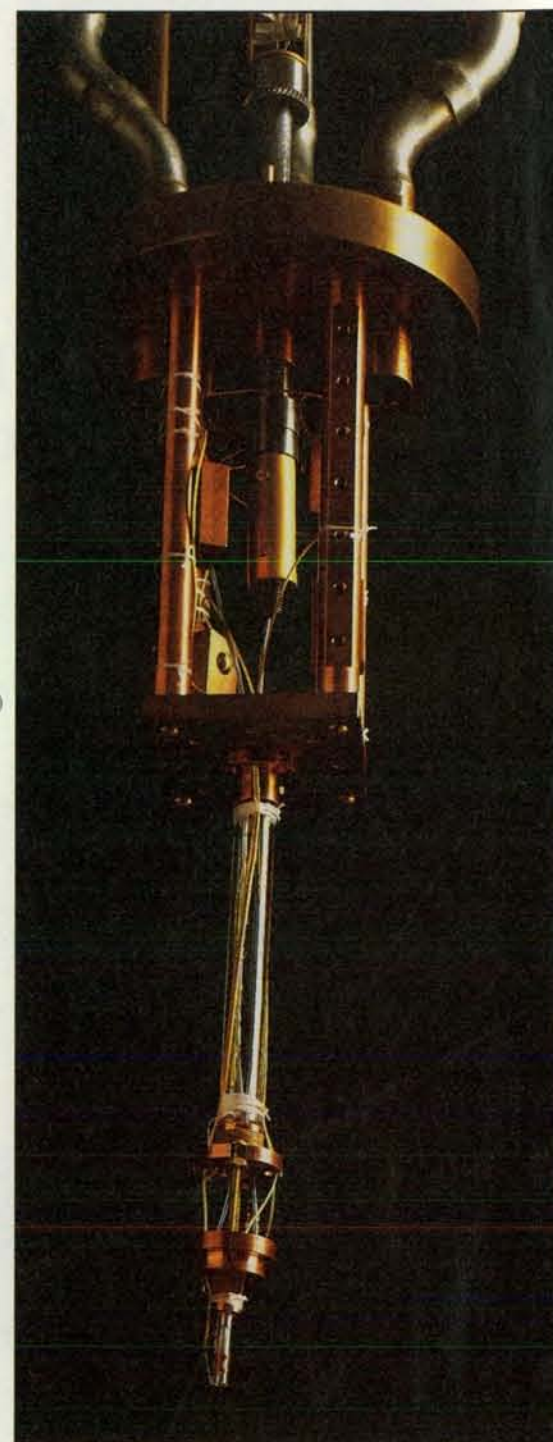
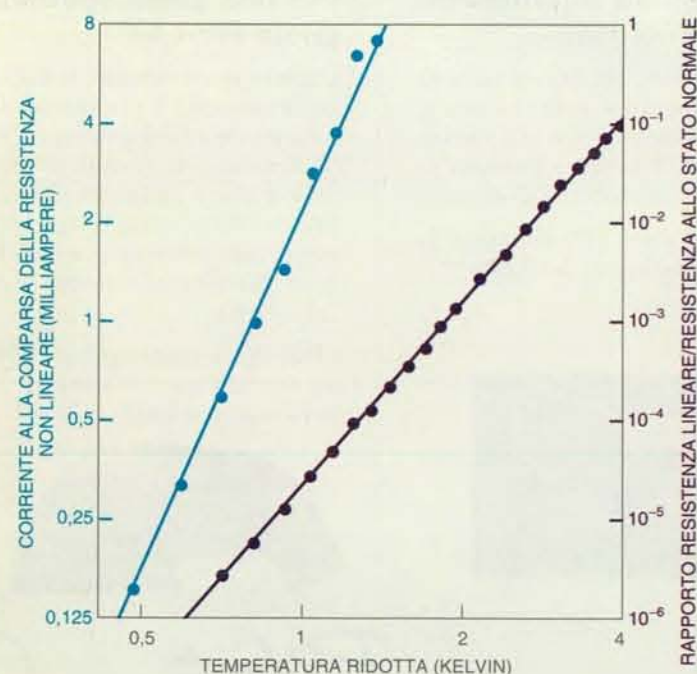
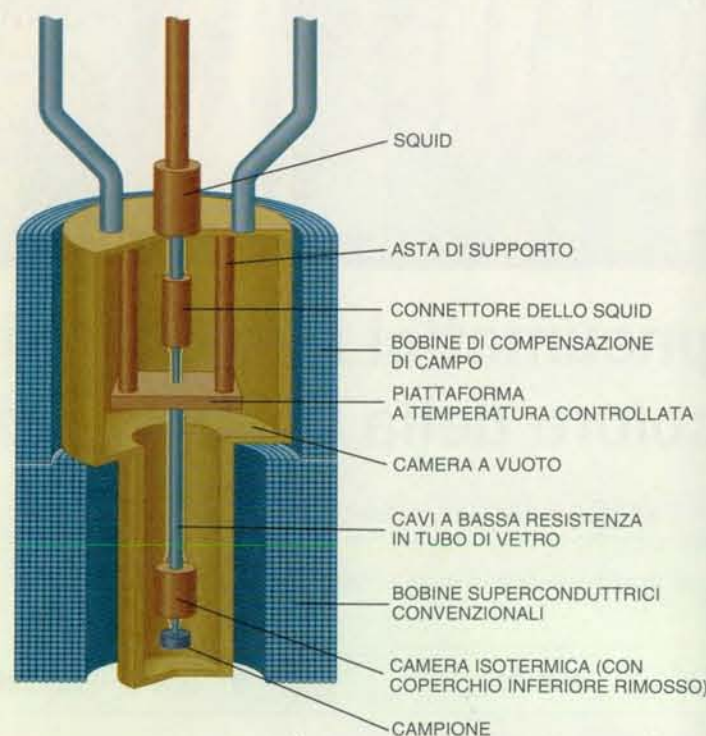
La domanda non è banale. I superconduttori previsti per uso tecnologico devono inevitabilmente essere «sporchi», dato che i ricercatori che lavorano sui superconduttori convenzionali producono materiali con difetti su misura. In generale più un superconduttore è sporco, più corrente può trasportare. La presenza di imperfezioni è auspicabile perché esse ancorano i vortici e impediscono loro di spostarsi sotto l'azione della forza di Magnus. Le linee di vortice risiedono preferenzialmente nei siti di ancoraggio del reticolo cristallino poiché in questo modo abbassano la propria energia.

Stati di un solido di vortici

Un superconduttore in un campo magnetico «congela» in un solido in due modi diversi. Se il materiale è pulito, le linee di vortice si ordinano in una disposizione triangolare regolare, un reticolo di vortici. Se la sostanza presenta molti difetti o impurezze le linee si dispongono in modo disordinato formando un cosiddetto vetro di vortici.

RETICOLO DI VORTICI

VETRO DI VORTICI

Il picovoltmetro SQUID

Questo dispositivo serve per analizzare i diversi stati di vortici. Le parti essenziali del picovoltmetro sono mostrate qui sotto a sinistra. Le bobine di superconduttore convenzionale e di compensazione del campo sono in grado di fornire un campo magnetico fino a sette tesla. La camera isotermica mantiene costante la temperatura del campione, a meno di qualche millikelvin. Cavi a bassa resistenza che passano attraverso un tubicino di vetro connettono il campione allo SQUID, il quale misura piccoli cambiamenti delle grandezze elettriche. Nella fotografia a destra le bobine, la camera a vuoto e la camera isotermica sono state rimosse per chiarezza. Misurazioni condotte con questo dispositivo hanno confermato il modello del vetro di vortici. In un esperimento i cui risultati sono riportati nel grafico, si sono misurate la corrente (*in blu*) e la resistenza (*in viola*) in una regione dove le proprietà elettriche del campione sono non lineari. I dati si dispongono in una linea retta, come previsto dalla teoria. La temperatura ridotta è data dalla differenza fra la temperatura del campione e quella alla quale si manifesta la fase superconduttiva di vetro di vortici.

La situazione è analoga a quella di una pallina che rotola su una tavola nella quale sono state praticate piccole buche. L'esperienza insegna che la pallina finisce per fermarsi in una delle buche, dove la sua energia potenziale gravitazionale è minima.

Il fenomeno dell'ancoraggio ha un effetto caratteristico sul solido di vortici in un superconduttore: esso distrugge il reticolo regolare che altrimenti si formerebbe in un materiale ideale puro. In altre parole impedisce al materiale di condensare in un perfetto solido di vortici in presenza di un intenso campo magnetico. Alla fase che invece si forma è stato ora assegnato il nome di «vetro di vortici»: il termine risulta appropriato, in quanto la posizione dei vortici ha un andamento disordinato e irregolare simile a quello assunto dalle molecole di un materiale vetroso.

Quando fu proposta per la prima volta, nel 1989, l'ipotesi del vetro di vortici non trovò grande consenso. Altre descrizioni, come per esempio quelle che trattano le linee di vortice come particelle singole, possono spiegare anch'esse il comportamento osservato nei superconduttori ad alta temperatura. Il modello del vetro di vortici ha fornito comunque alcune previsioni verificabili. Esso postulava per esempio che, data una sufficiente concentrazione di difetti che possono fungere da siti di ancoraggio, il liquido di vortici si congelasse gradualmente in un vetro. Questo comportamento contrasta con quanto avviene nei materiali puri, nei quali il liquido di vortici solidifica abbastanza bruscamente e in modo isteretico. Il modello del vetro di vortici descriveva anche il comportamento della resistività in funzione della temperatura, della corrente e del campo magnetico.

Una chiara verifica del modello del vetro di vortici si ebbe solo quando si riuscirono a compiere misurazioni di trasporto estremamente accurate, di un tipo non usualmente utilizzato nel campo della superconduttività. In particolare è stato progettato un apparecchio che potesse misurare la differenza di potenziale ai capi di un superconduttore ad alta temperatura con una risoluzione dell'ordine dei picovolt (10^{-12} volt), che precedentemente non era ottenibile. Il picovoltmetro usava un dispositivo superconduttore a interferenza quantistica o SQUID; questi dispositivi fondano il proprio funzionamento su effetti quantistici per misurare minuscole variazioni di corrente e differenza di potenziale. Grazie allo SQUID, il picovoltmetro aveva una sensibilità circa un milione di volte più grande di quella di un voltmetro ordinario. La risoluzione era dunque sufficientemente elevata per poter confermare o confutare la teoria del vetro di vortici.

Il principio di funzionamento del picovoltmetro in sé è molto semplice. Il campione viene posto in un contenitore coibentato che riesce a mantenere la temperatura costante a meno di pochi millikelvin. Bobine superconduttrici avvolgono il contenitore e producono il campo magnetico uniforme che deve essere applicato al campione. La corrente viene inviata al campione mentre lo SQUID ne misura la resistenza. Lo SQUID e i magneti sono superconduttori convenzionali a bassa temperatura, un esempio di come la vecchia tecnologia ci aiuti a studiare e comprendere la nuova.

Il raffreddamento e il riscaldamento di un cristallo molto puro del superconduttore YBCO in un campo magnetico producono diagrammi di resistenza che non si sovrappongono perfettamente. La misurazione indica che il reticolo di vortici fonde in maniera repentina. Il liquido di vortici può venire leggermente sottoraffreddato prima di congelare, analogamente a quanto avviene per l'acqua distillata.

L'apparecchiatura descritta ha confermato inequivocabilmente l'esattezza delle previsioni del modello del vetro di vortici. Le resistenze e le correnti misurate erano in perfetto accordo con quelle previste dal modello, e scendevano uniformemente a zero al diminuire della temperatura verso il punto di congelamento del liquido di vortici. Questo comportamento è molto diverso da quello mostrato da cristalli molto puri: in essi la transizione di fase è improvvisa e isteretica. Le osservazioni dimostrano l'importanza del disordine indotto dai siti di ancoraggio - il ruolo dello «sporco», per così dire - nel cambiamento della dinamica della transizione. Invece che in un solido di vortici, il liquido di vortici nel cristallo disordinato congela in un vetro di vortici.

I superconduttori ad alta temperatura si sono rivelati un fantastico banco di prova per le nostre conoscenze nel campo della superconduttività di tipo II. Per esempio si può affermare che il vetro di vortici esiste anche nei superconduttori convenzionali, sebbene questo stato della materia sia in essi difficile da osservare. Tuttavia resta da vedere se la nostra conoscenza possa essere tradotta efficacemente in applicazioni. Si sta al momento indagando attivamente per trovare difetti che possano ancorare i vortici nel modo più efficace. Molto è già stato fatto per produrre cavi superconduttori e per incrementare le densità di corrente trasportate. Le nostre conoscenze a livello microscopico sui vari stati di vortici possono solo aiutarci a produrre materiali migliori per rendere possibili quelle applicazioni che vengono attese con tanta impazienza.

BIBLIOGRAFIA

PARKS R. D. (a cura), *Superconductivity*, Marcel Dekker, 1969.

TINKHAM MICHAEL, *Introduction to Superconductivity*, Robert E. Krieger Publishing, 1980.

BISHOP D. J., GAMMEL P. L., HUSE D. A. e MURRAY C. A., *Magnetic Flux-Line Lattices and Vortices in the Copper Oxide Superconductors* in «Science», 255, pp. 165-172, 10 gennaio 1992.

HUSE D. A., FISHER MATTHEW P. A. e FISHER DANIEL S., *Are Superconductors Really Superconducting?* in «Nature», 358, n. 6387, 13 agosto 1992.

Le fratture del femore: un'epidemia ortopedica

In tutti i paesi occidentali questo tipo di incidente è diventato un evento sempre più frequente tra la popolazione ultrasessantenne, con gravi conseguenze psicologiche e sociali e con costi sanitari sempre crescenti

di Paolo Della Torre, Piero Petrini ed Emanuele Rinonapoli

Le fratture dell'estremo superiore del femore sono per la loro frequenza e per la loro gravità uno dei problemi di traumatologia più importanti e studiati. Nel loro complesso rappresentano circa il 27-30 per cento di tutte le fratture e, nella sola Italia, il loro numero globale è stimato intorno ai 30 000 casi all'anno. Sono appannaggio quasi esclusivo dei soggetti al di sopra dei 60-65 anni, aumentano con l'età e sono molto più comuni nelle donne che negli uomini (il rapporto è di circa 2-3:1).

Questo tipo di fratture rappresenta quasi sempre la conseguenza più grave di un evento molto comune in età avanzata: la caduta accidentale.

Sia in ambiente domestico, sia all'esterno della propria abitazione, le cadute accidentali sono uno dei problemi più gravi dell'età avanzata per due ordini di motivi: la grande frequenza con cui si verificano e le conseguenze spesso molto serie che esse comportano.

Per quanto riguarda la frequenza, bisogna tenere presente che secondo molti studi epidemiologici si ritiene che circa il 40 per cento delle donne e il 20 per cento degli uomini oltre i 65 anni di età abbiano una storia di cadute e che la metà degli ultraottantenni sia incorsa in una caduta importante nell'anno precedente l'indagine epidemiologica. Le cause delle cadute sono molteplici e i fattori di rischio sono di tipo ambientale e individuale.

Tra i primi spesso risulta determinante l'ambiente domestico, in particolare le scale male illuminate con gli spigoli dei gradini in ombra, i pavimenti bagnati, l'uso di pantofole o zoccoli, i tappeti non fissi, i cavi elettrici mobili e così via.

I fattori di rischio individuali sono soprattutto rappresentati dall'età e da varie condizioni di salute come, per esempio, difetti della vista, alterazioni della validità degli arti inferiori, disturbi dell'equilibrio, episodi di transitoria scarsa irrorazione sanguigna cerebrale, uso di alcool e assunzione di farmaci come sedativi, antidepressivi, anti-ipertensivi, diuretici, che possono far rallentare i riflessi di reazione o indurre brevi perdite di coscienza per un calo della pressione del sangue nel passare dalla posizione sdraiata o seduta a quella eretta.

Naturalmente spesso tali fattori di rischio risultano associati, come nelle istituzioni per anziani dove infatti la frequenza delle cadute è in genere più elevata, il che è comprensibile se si tiene presente che in genere gli ospiti delle case di riposo sono meno «sani» e quindi più medicalizzati e vivono in un ambiente che, almeno all'inizio, presenta caratteristiche e «ostacoli» non familiari.

Per fortuna non tutte le cadute causano fratture e anzi si ritiene che la percentuale di cadute che hanno come esito una frattura oscilli tra l'uno e il sei per cento a seconda delle caratteristiche dei gruppi presi in esame.

Tuttavia una caduta in un anziano ha sempre rilevanti conseguenze psicoemotive, soprattutto per il timore di altre cadute che porta spesso a chiudersi entro le mura domestiche e a uno stato di depressione e di dipendenza. Non di rado tali timori sono vissuti anche dalla famiglia che può diventare iperprotettiva e, per paura di nuovi incidenti, decidere di ricoverare l'anziano in un istituto.

Tutto questo si traduce in una riduzione della mobilità e dell'autonomia a causa della mancanza di sostegni e di sollecitazioni psichiche.

Se poi l'anziano vive da solo, le conseguenze sono ancora più gravi: per mancanza di tempestivi soccorsi, il paziente può restare a lungo per terra con danni sia fisici sia psicologici.

Le fratture di gran lunga più frequenti collegabili alle cadute accidentali sono quelle del polso e del femore, che tuttavia presentano caratteristiche notevolmente diverse tra di loro.

Le prime, molto comuni, sono fratture dell'estremità del radio e sono chiamate anche fratture di Colles dal nome del chirurgo che per primo le descrisse. Esse sono comuni a tutte le età e sono legate al riflesso istintivo di appoggio sulla mano nel caso di caduta. Nell'anziano queste fratture aumentano di incidenza, in rapporto alla progressiva maggiore fragilità ossea, fino a circa 75 anni; dopo questa età esse si riducono nettamente, probabilmente a causa del venir meno del riflesso di appoggio istintivo, ciò che, come vedremo, comporta invece un aumento delle fratture del femore. Le fratture della regione del polso sono in genere a prognosi buona e comportano, nella grande maggioranza dei casi, un trattamento di sola immobilizzazione per circa 30-40 giorni.

Le fratture dell'estremo superiore del femore sono per lo più imputabili a un meccanismo indiretto con urto della regione laterale dell'anca. Come vedremo più avanti, tali lesioni sono certamente collegabili a un insieme di fattori connessi ai processi degenerativi della senescenza, tra i quali ha grande rilievo la perdità della qualità meccanica del tessuto osseo.

È consuetudine classificare le fratture del femore a seconda del decorso della rima di frattura: si hanno fratture del collo del femore e fratture pertrocanteriche.

FRATTURE DEL FEMORE ALL'ANNO (PER 10 000 ABITANTI)
150
100
50
DONNE
UOMINI
50 60 70 80
ETÀ

Le fratture dell'estremo superiore del femore sono rare al di sotto dei 50 anni. Sopra i 60 anni, invece, il tasso d'incidenza aumenta in funzione dell'età, soprattutto nel sesso femminile.

Il decorso della rima di frattura del femore condiziona il tipo di intervento e il ripristino della funzionalità. In queste sequenze d'immagini sono rappresentati rispettivamente il disegno anatomico della frattura, il suo aspetto radiografico e il trattamento consigliato. In alto, una frattura del collo del femore, un trauma in cui le sfavorevoli condizioni meccaniche e il frequente danno vascolare possono provocare, in un'alta percentuale di pazienti, mancata consolidazione della frattura e necrosi ossea. Nei soggetti molto anziani la sostituzione della testa femorale con una protesi consente un'immediata ripresa della deambulazione. La frattura pertrocanterica del femore considerata nella sequenza in basso non presenta, invece, importanti problemi di consolidazione. Lo schema terapeutico prevede un intervento per ricomporre la frattura e per fissarla con un mezzo metallico di sintesi, che consente di evitare l'immobilizzazione esterna e di ripristinare rapidamente la mobilità.

QUINQUENNIO	POPOLAZIONE TOTALE (MEDIA)	ULTRA--SESSANTENNI	FRATTURE DEL FEMORE (TOTALE)	FRATTURE DEL FEMORE (ALL'ANNO)
1975-1979	138 125	24 454	223	44,6
1987-1991	148 621	30 651 (+ 25%)	467	93,4 (+109%)

Questa tabella evidenzia l'evoluzione della popolazione del comune di Perugia tra il quinquennio 1975-1979 e il quinquennio 1987-1991. Gli autori hanno preso in considerazione nel loro studio la popolazione ultrasessantenne, il gruppo più a rischio per le fratture del femore. A un aumento di questo segmento della popolazione del 25 per cento è corrisposto un incremento delle fratture del femore del 109 per cento.

Tali fratture, che rappresentano una delle più comuni e pericolose situazioni cliniche dell'età avanzata, possono essere di notevolissima gravità. Fino a qualche decennio fa esse rappresentavano una delle principali cause di morte nella vecchiaia e ciò non tanto per l'azione diretta del trauma in sé, quanto per l'instaurarsi di complicazioni polmonari, circolatorie, renali e locali cutanee, che quasi inevitabilmente si verificano in pazienti, talora molto anziani, costretti alla immobilità.

Un notevole passo in avanti nel settore della chirurgia ortopedica è stato compiuto grazie all'avvento della osteosintesi, una tecnica che, fissando chirurgicamente i frammenti ossei, non solo realizza le condizioni locali più favorevoli al processo di consolidazione, ma altresì evita agli infortunati l'immobilità, che è la causa principale delle complicazioni. Più recentemente la possibilità di sostituire con protesi artificiali l'estremità superiore del femore (la cosiddetta testa femorale) ha consentito un ulteriore progresso per quei non rari tipi di frattura del collo del femore in cui esiste un'alta probabilità che si verifichino serie complicazioni a distanza di tempo. Esse sono rappresentate da mancata consolidazione della frattura o da cedimento del tessuto osseo della testa femorale a causa di una necrosi determinata, a sua volta, da un irreversibile danno circolatorio legato al trauma stesso.

Attualmente, dunque, tutte le fratture dell'estremo superiore del femore vengono trattate chirurgicamente: questo tipo di intervento consente sempre una precoce mobilizzazione dei pazienti e in parecchi casi anche una veloce ripresa della deambulazione, permettendo quindi di prevenire le complicanze più pericolose.

Con le moderne tecniche di anestesia generale o più spesso regionale, né l'età avanzata, né le cattive condizioni generali sono, di norma, controindicazioni per l'operazione che, nel complesso, è poco traumatizzante.

In genere lo schema terapeutico adottato per le fratture del collo del femore prevede, fino a 60-65 anni di età del paziente, la osteosintesi con viti o chiodi per conservare la struttura anatomica; in pazienti oltre i 65 anni, constatato l'alto rischio che tali fratture non consolidino (dando origine alla cosiddetta pseudoartrosi) o che evolvano verso una necrosi della testa femorale per carenza di irrorazione sanguigna, non si tenta nemmeno di ottenere la consolidazione, ma si sostituisce la testa femorale separata con una testa artificiale, chiamata protesi cefalica, che può essere semplice o a doppio snodo per ridurre l'attrito tra la testa artificiale metallica e il rivestimento cartilagineo dell'acetabolo.

Per le fratture della regione trocanterica si esegue sempre la osteosintesi con placche, viti-placche o chiodi endomidollari, avendo come obiettivo principale quello di ripristinare la massima stabilità meccanica con la minima invasività chirurgica.

L'esito finale di queste lesioni, come abbiamo detto, è molto migliorato con l'avvento delle tecniche chirurgiche, e accurati studi comparati hanno accertato che la mortalità entro sei mesi dal trauma è passata da un 40-50 per cento all'attuale 10-20 per cento.

Per quanto riguarda le complicazioni locali o generali, esse sono ancora abbastanza frequenti e si aggirano intorno al 20-30 per cento; molte sono lievi e transitorie e quindi facilmente controllabili; altre come le complicanze respiratorie e vascolari e i decubiti sono di maggiore impegno terapeutico e responsabili di uno scadimento della qualità della vita.

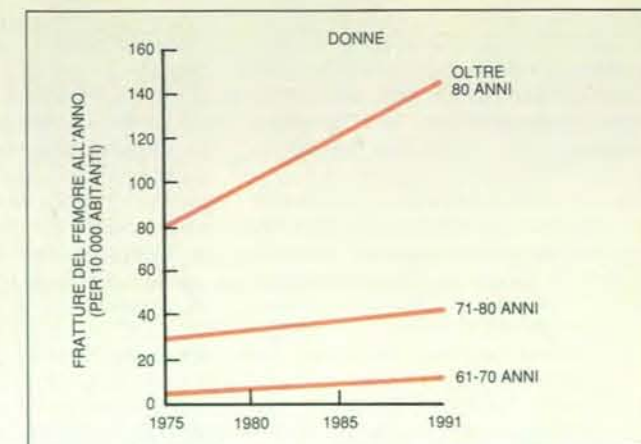
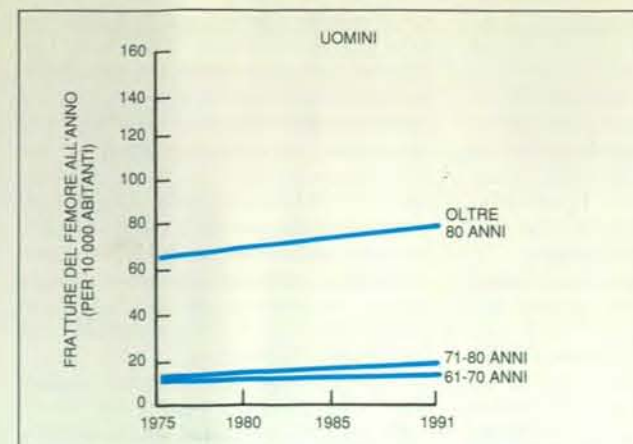
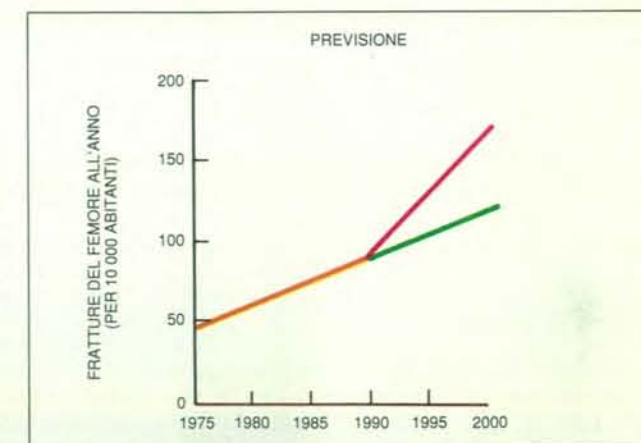
A proposito della qualità del recupero e della vita di relazione dopo il trauma e l'intervento, è interessante notare che circa l'80-85 per cento dei pazienti riesce a riprendere le normali attività e una vita di relazione simile a quella precedente solo dopo alcuni mesi dall'incidente.

I fattori prognostici che influenzano la qualità del risultato finale sono essenzialmente l'età e soprattutto lo stato fisico generale in rapporto alla presenza o meno di altre malattie. È stato infatti ampiamente dimostrato che l'età di per se stessa non rappresenta un sicuro elemento negativo, mentre la presenza di condizioni generali scadenti o di malattie di altri organi e apparati (ovviamente più probabili con l'avanzare dell'età) costituisce un sicuro fattore di rischio. Esso è tanto maggiore quanto più serie sono le affezioni presenti e incide sia sulla possibilità di sopravvivenza sia sulla qualità del recupero. Va sottolineato che il più decisivo fattore di guarigione resta pur sempre la volontà di guarire che, nel caso di una frattura di femore operata, è messa decisamente alla prova. Purtroppo non è raro il caso di soggetti che, pur venendo tempestivamente operati, non approfittano della possibilità di riprendere a camminare e si trattengono a letto soprattutto per mancanza di volontà o di stimoli, correndo così il rischio di gravi complicazioni e ritardando o compromettendo la ripresa funzionale.

Circa la tempestività dell'intervento chirurgico si è molto discusso tra gli addetti ai lavori se sia più vantaggiosa una soluzione chirurgica immediata o ritardata di alcuni giorni. La prima comporta una incompleta conoscenza dello stato di salute del paziente, ma garantisce una precocissima ripresa di motilità dell'arto leso e spesso della deambulazione. La seconda prevede l'intervento dopo una attenta valutazione e preparazione del traumatizzato, ma aggrava quei fenomeni come i decubiti e lo scadimento dell'umore del paziente che sono proporzionali alla durata del periodo di allettamento preoperatorio. È verosimile che la condotta più corretta sia un giusto equilibrio tra le due esigenze di conoscenza clinica del paziente e precocità dell'intervento senza ulteriori ritardi.

È chiaro, comunque, che tale lesione, per la sua alta incidenza, per la sua gravità, per la tipologia del paziente affetto, per la necessità di una terapia sempre chirurgica in ambiente ospedaliero, per i lunghi tempi di recupero e di assistenza fisioterapica, comporta elevati costi umani, sociali e organizzativi.

A proposito dei costi sociali va sottolineato che essi devono includere la spesa per la degenza necessaria al trattamento e quella per la lunga terapia riabilitativa e la invalidità residua che talora rende indispensabile un'assistenza domiciliare. Per l'Italia tali costi sono stati approssimativamente stimati in 730 miliardi l'anno per le degenze ospedaliere e in 364 000 giornate lavorative perse.

È nota la diversa incidenza delle fratture prossimali del femore in relazione alla diversa composizione etnica, sociale e geografica delle popolazioni prese in esame. In particolare essa risulta più alta nelle popolazioni caucasiche rispetto a quelle di origine africana e nelle regioni nordiche rispetto a quelle della fascia mediterranea o tropicale.

I due grafici in alto, costruiti in base ai dati riportati nelle tabelle di pagina 44, rappresentano l'andamento dell'incidenza delle fratture del femore nei due sessi e per decenni di età nel comune di Perugia. L'aumento è evidente e statisticamente significativo nelle donne al di sopra dei 70 anni. Qui a fianco la previsione per l'anno 2001 del numero di fratture del femore attese nel comune di Perugia in base alle proiezioni demografiche che sono state elaborate dall'ISTAT. L'andamento che si avrebbe se si riuscisse a stabilizzare l'incidenza intorno ai valori attuali (incremento del 22 per cento) è in verde; in viola, quello che si avrebbe se proseguisse l'attuale tendenza all'aumento delle fratture (incremento dell'88 per cento).

Negli ultimi quarant'anni è stato inoltre notato, in varie regioni del mondo occidentale, un sensibile aumento di tali lesioni anche se con notevoli variazioni in rapporto ai vari gruppi etnici e sociali presi in considerazione.

Il problema fondamentale è se tale incremento debba essere riferito al solo effetto delle variazioni demografiche, e cioè all'aumento del numero di anziani (ultrasessantenni), in particolare donne, nelle popolazioni prese in esame, o se vi siano altri fattori eziopatogenetici in gioco, ancorché sconosciuti, che determinino un aumento dell'incidenza specifica per ogni gruppo di età. Recentemente alcuni studi condotti in comunità più o meno grandi dell'Europa settentrionale hanno permesso di accertare che, negli ultimi 20-25 anni, presso le località prese in esame, l'aumento del tasso di incidenza di tali fratture è andato ben al di là di quello atteso per la sola diversa distribuzione di età della popolazione: esso si è dimostrato in costante aumento anche all'interno di classi di età omogenee, pur con sensibili differenze tra le varie zone geografiche.

Per quanto riguarda il nostro paese, abbiamo avuto la possibilità di verificare l'incidenza assoluta delle fratture del femore e le sue variazioni, in rapporto all'età e al sesso, in una popolazione italiana omogenea e tecnicamente ben esaminabile per il periodo relativo agli ultimi 16 anni. La premessa per la corretta realizzazione di uno studio epidemiologico su questo tipo di fratture è, infatti, la conoscenza di due parametri fondamentali: *a*) la composizione anagrafica e il suo variare nel tempo (per quinquenni o decenni di età) di una popolazione in esame; *b*) il numero di fratture del femore che, storicamente, si sono verificate in quella stessa popolazione.

Nel comune di Perugia, tutte le fratture del femore vengono trattate presso un solo reparto ospedaliero: la Clinica ortopedica dell'Università di Perugia. Il numero di fratture verificate è stato quindi calcolato sulla base dei registri statistici e dei dati di archivio di tale Istituto clinico. È teoricamente possibile che alcuni residenti nel comune di Perugia siano stati trattati in altri ospedali e che quindi sfuggano al nostro rilievo, ma certamente sono casi rari che non possono avere influenza pratica sui risultati di tale studio. Nel computo globale delle fratture dei soli residenti nel comune di Perugia, esse sono state divise in «fratture del collo del femore» (più interne), e «fratture pertrocanteriche» (più esterne), che, come abbiamo visto, hanno una diversa natura e schema terapeutico. Abbiamo inoltre escluso dalla nostra indagine i pazienti di età inferiore a 60 anni, le cui fratture del femore sono di solito imputabili a traumi «di alta energia» e sono perciò del tutto differenti dalle ordinarie fratture dell'anziano, che generalmente sono dovute a semplici cadute; le fratture patologiche, cioè quelle la cui causa principale è rappresentata da una lesione ossea dovuta a una patologia specifica e locale (come le metastasi ossee); le fratture parcellari con interessamento della sola testa femorale o del solo gran trocantere; le rare fratture la cui eziologia era riconducibile a un «trauma maggiore», come un incidente stradale, e che quindi non potevano essere considerate nell'indagine epidemiologica.

Per realizzare un significativo confronto a distanza di circa 12 anni e per avere una base statistica più ampia abbiamo esaminato il numero di fratture verificatesi nel quinquennio 1975-1979 e quello del quinquennio 1987-1991. Va premesso che nel primo quinquennio esaminato il comune di Perugia aveva in media 138 125 abitanti con 24 454 ultrasessantenni (17,7 per cento). Nel secondo quinquennio la popolazione media è salita a 148 621 abitanti, dei quali 30 651 erano ultrasessantenni (20,6 per cento). L'incremento della popolazione a rischio è stato perciò del 25 per cento.

Il numero di fratture dell'estremo prossimale del femore, in pazienti con

età superiore a 60 anni residenti nel comune di Perugia, è aumentato da 44,6 all'anno, di media, negli anni 1975-1979 a 93,4 all'anno negli anni 1987-1991. Quindi, in un intervallo di 12 anni si è avuto un incremento globale del 109 per cento. Conteggiando separatamente le fratture del collo del femore e le fratture trocanteriche si è notato un aumento non uniforme tra questi due gruppi. Infatti le fratture del collo femorale hanno avuto un incremento del 106 per cento mentre quelle trocanteriche del 145 per cento.

La distribuzione di tali fratture e la loro incidenza specifica per sesso e decenni di età, verificatasi nel primo e secondo quinquennio sono riportate in dettaglio nella tabella a pagina 42. In tutti i sei gruppi considerati si è avuto un incremento di incidenza, ma si sono potute notare sensibili differenze. Nella popolazione maschile si è infatti osservato un incremento più omogeneo e meno marcato; lo stesso fenomeno si è riscontrato nella popolazione femminile tra i 60 e gli 80 anni. Un incremento marcato si è verificato invece per le donne al di sopra di 80 anni, ove l'incidenza è passata da 81/10 000 (e cioè un caso di frattura ogni 123 donne all'anno) a 146/10 000 (un caso ogni 68 donne all'anno).

Sulla base delle proiezioni demografiche elaborate dall'ISTAT per la regione Umbria e riferite all'anno 2001, si può prevedere il numero di fratture dell'estremo prossimale del femore attese per quell'anno secondo due ipotesi. Se il numero di fratture aumentasse in base al solo incremento demografico delle classi di età a rischio, ma il tasso di incidenza si mantenesse stabile rispetto all'attuale, avremmo a Perugia 125 fratture all'anno (con un incremento del 34 per cento). Se invece il numero di fratture seguisse le variazioni demografiche e inoltre l'incidenza aumentasse proseguendo l'attuale tendenza, si verificherebbero allora 176 fratture all'anno (con un incremento dell'88 per cento).

I dati riferiti al comune di Perugia - che possiamo considerare un possibile, significativo campione per l'Italia - mostrano una tendenza all'aumento abbastanza omogenea rispetto a quella già notata in altri paesi europei per lo stesso periodo di riferimento, come evidenzia il confronto dei dati riportato nella pagina a fronte. Infatti il tasso di aumento è quasi sempre maggiore per le donne, soprattutto per le più anziane. Tuttavia i valori sia di partenza sia di arrivo sono sempre più bassi a conferma di un fenomeno già noto, e cioè che il fattore di rischio è più basso alle nostre latitudini rispetto all'Europa settentrionale.

L'incremento nell'incidenza di queste fratture legate a traumi minori può essere attribuito a vari fattori che possono spiegare perché le cadute negli anziani siano più frequenti e le loro conseguenze siano più gravi.

Innanzitutto va sottolineato che tra i motivi che possono favorire la caduta in età avanzata ve ne sono alcuni di ordine sanitario collegabili a una serie di concomitanti patologie a vari organi e distretti. In particolare sono ritenute responsabili le affezioni del sistema nervoso centrale come l'ictus, e non rare crisi epilettiche minori, il morbo di Parkinson per la nota difficoltà alla deambulazione che esso comporta, i deficit transitori di irrorazione cerebrale che provocano un'improvvisa perdita di tono della muscolatura degli arti con conseguente caduta. Anche le affezioni del sistema cardiovascolare, come i disturbi transitori del ritmo cardiaco e le crisi ipotensive posturali, possono causare cadute, così come la riduzione della capacità visiva per patologie oculari quali la cataratta, il glaucoma, le retinopatie, i disturbi dell'equilibrio riferibili a patologie vestibolari o a perdita dei riflessi propriocettivi cervicali e lombo-sacrali e le malattie degenerative osteoarticolari come le comunissime osteoartrosi di anca e ginocchio.

È ipotizzabile che i recenti progressi medici e sociali abbiano determinato un aumento della sopravvivenza degli anziani con patologie croniche e allungato le loro aspettative di vita, ma che essi siano in qualche modo meno «sani» rispetto a quelli di alcuni decenni fa. Anche fattori esterni possono essere incolpati di favorire le cadute degli anziani; in particolare è stata segnalata l'associazione tra l'aumento delle cadute e il recente incremento, nelle classi di età più avanzata, del consumo di farmaci che possono causare vertigini e di alcool.

Neppure i fattori ambientali vanno trascurati.

È stato notato che un terzo delle cadute si verifica entro le mura domestiche e che la responsabilità di un terzo di queste è attribuibile alle scale.

Tali rilievi tuttavia, se possono spiegare alcune rilevanti differenze geografiche nell'incidenza delle cadute e delle fratture (nel contesto abitativo anglosassone, per esempio, le cadute dalle scale sono tipiche), non possono spiegare l'incremento recente di tale fattore di rischio in Italia.

Un aspetto di grande importanza nella patogenesi delle fratture è la qualità dell'osso senile. È stato dimostrato che l'attività fisica, la dieta, eventuali interventi di ovariectomia e di gastroresezione, l'esposizione solare (attraverso la sintesi di vitamina D), l'uso di alcool e di tabacco influiscono sulla struttura del tessuto osseo. Il ruolo svolto dall'osteoporosi nell'aumentato tasso di fratture del femore è dimostrato dal fatto che si è osservato un aumento specifico delle fratture della regione trocanterica (quella più colpita dall'osteoporosi) rispetto a quelle del collo del femore. Inoltre l'incremento di tali fratture verificatosi negli ultimi decenni è stato soprattutto a carico delle classi di età più avanzate e del sesso femminile, ove il ruolo causale dell'osteoporosi è indiscusso. Infine alcuni ricercatori hanno rilevato una sensibile differenza di incidenza tra le aree urbane e quelle rurali, ove l'attività fisica, l'esposizione al sole e la dieta costituiscono probabilmente fattori di protezione.

Va notato a questo proposito che la diffusa urbanizzazione verificatasi in tutto l'Occidente (e anche nell'ambito del comune di Perugia) nella seconda metà di questo secolo può fornire una chiave interpretativa di tale fenomeno epidemiologico.

Tuttavia le ragioni concrete di un incremento così rilevante e concorde del tasso di incidenza per età e sesso delle fratture prossimali del femore restano oscure. È probabile che siano contemporaneamente in gioco più elementi e sono necessari altri studi e analisi multifattoriali per individuarli.

Il dato epidemiologico resta comunque assai allarmante e ha indotto a parlare di una vera e propria «epidemia delle fratture prossimali del femore». Solo per fare un esempio già nel 1985 i pazienti con tali lesioni occupavano il 20 per cento dei posti letto ortopedici in Gran Bretagna.

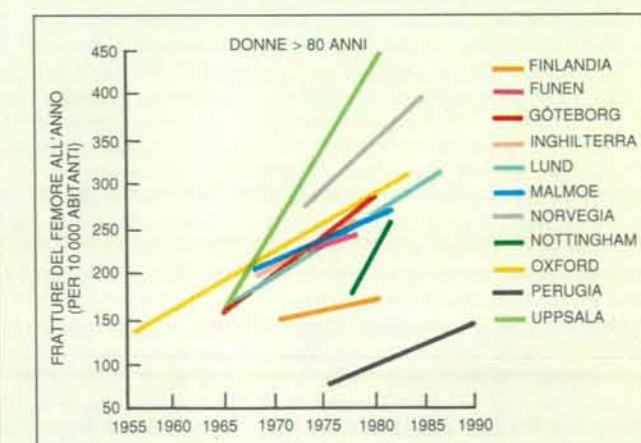
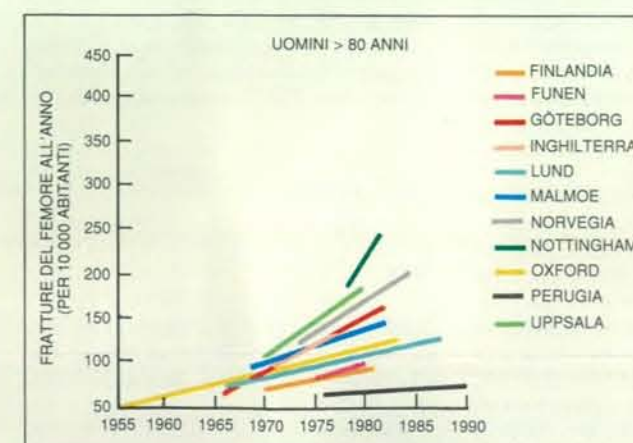
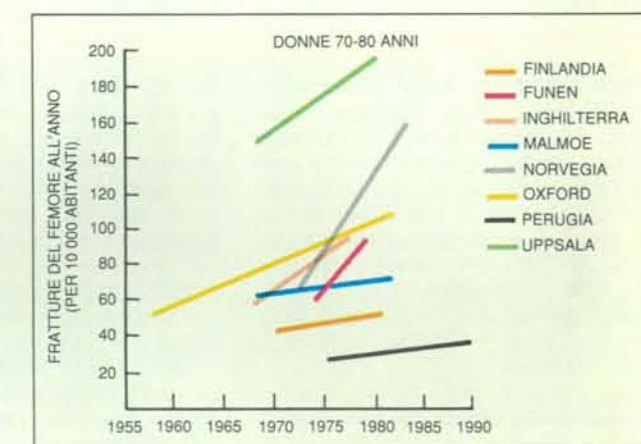
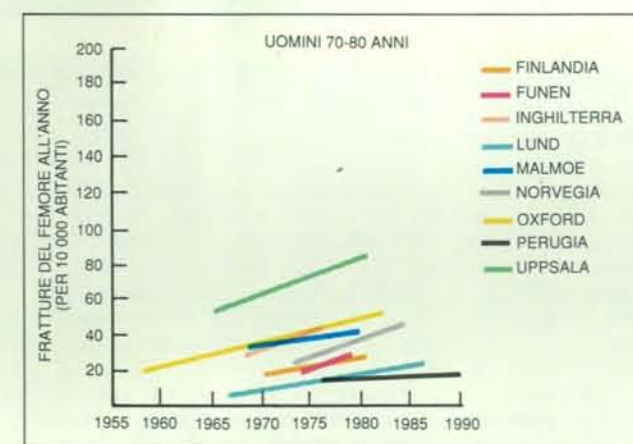
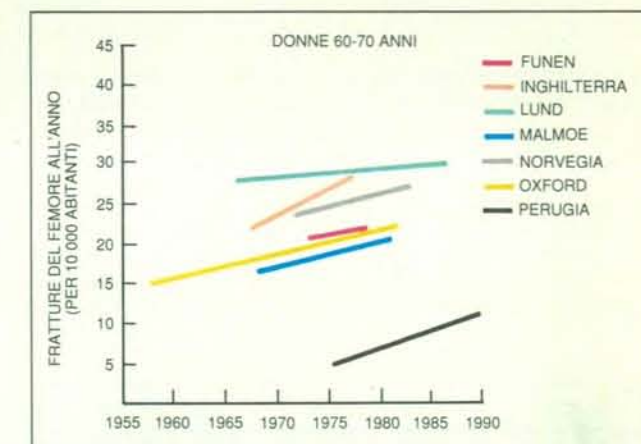
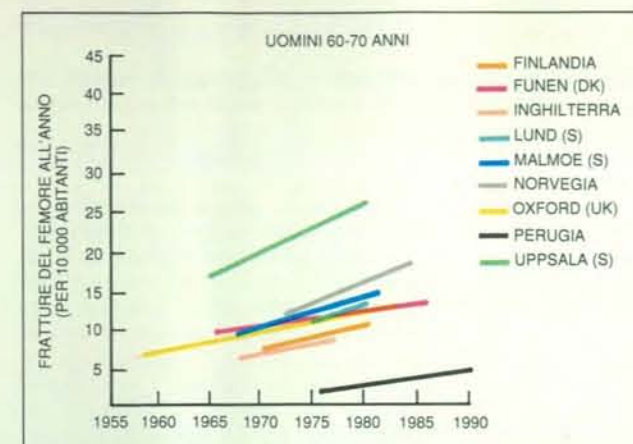
Questo aspetto di patologia traumatica geriatrica deve essere adeguatamente preso in considerazione nella pianificazione socio-sanitaria nazionale, che giustamente da alcuni anni si preoccupa delle conseguenze del progressivo invecchiamento demografico della popolazione italiana.

UOMINI

ETÀ	1975-1979 SOGGETTI	1975-1979 FRATTURE DEL FEMORE (TOTALE)	1975-1979 FRATTURE DEL FEMORE (/10 000/ANNO)	1987-1991 SOGGETTI	1987-1991 FRATTURE DEL FEMORE (TOTALE)	1987-1991 FRATTURE DEL FEMORE (/10 000/ANNO)
61-70	6490	7	2,2	7251	17	4,7
71-80	3076	22	14,3	4380	39	17,8
>80	974	33	67,8	1153	46	79,8
TOTALE	10 540	62	11,8	12 784	102	15,9

DONNE

ETÀ	1975-1979 SOGGETTI	1975-1979 FRATTURE DEL FEMORE (TOTALE)	1975-1979 FRATTURE DEL FEMORE (/10 000/ANNO)	1987-1991 SOGGETTI	1987-1991 FRATTURE DEL FEMORE (TOTALE)	1987-1991 FRATTURE DEL FEMORE (/10 000/ANNO)
61-70	7633	20	5,2	8847	49	11,1
71-80	4378	64	29,2	6460	129	39,9
>80	1903	77	80,9	2560	187	146,1
TOTALE	13 914	161	23,1	17 867	365	40,9

Le tabelle riassumono, relativamente ai quinquenni considerati e ai residenti del comune di Perugia, la distribuzione per sesso e decenni di età delle fratture dell'estremo superiore del femore e l'incidenza annua sulla popolazione complessiva.

Le rilevazioni sull'incidenza della frattura del femore effettuate negli ultimi trent'anni in varie località dell'Europa settentrionale sono confrontate con i dati del comune di Perugia. Si è constatato che il tasso di aumento, espresso dall'inclinazione delle rette, è sempre maggiore nelle donne, soprattutto in quelle più anziane. Questa tendenza si manifesta anche nel comune di Perugia, ma i valori iniziali e finali sono più bassi, segno che il fattore di rischio nel nostro paese è minore.

Fattori che influenzano le cadute	Fattori che influenzano la qualità ossea
- Presenza di condizioni patologiche varie • affezioni del sistema nervoso centrale • affezioni del sistema cardio-vascolare • riduzione della capacità visiva • disturbi dell'equilibrio • malattie osteoarticolari - Uso di medicinali • psicofarmaci • antiipertensivi • diuretici - Consumo di alcool - Fattori ambientali	- Attività fisica - Dieta - Esposizione solare - Interventi di gastroresezione - Interventi di ovariectomia - Uso di alcool e tabacco

L'incremento dell'incidenza delle fratture del femore trova una spiegazione nel progressivo invecchiamento della popolazione e quindi nel numero sempre più elevato di anziani sottoposti a terapie e con una cattiva qualità del tessuto osseo.

La ricerca epidemiologica va finalizzata alla conoscenza dei fenomeni morbosi e del loro modificarsi nel tempo, ma con l'obiettivo finale di cogliere, se possibile, indicazioni utili per validi interventi preventivi. A proposito delle fratture del collo del femore va innanzi tutto contrastata la fatalistica concezione che le cadute, la osteoporosi senile, e le conseguenti fratture siano elementi inevitabili dell'età avanzata.

Un articolato programma di prevenzione deve avere come scopo quello di incidere sulle cause modificabili che portano alle fratture e cioè sulla riduzione dei fattori di rischio di cadute e sulla qualità del tessuto osseo degli anziani.

Un intervento di prevenzione delle cadute degli anziani non deve essere esclusivamente passivo, e anzi è stato segnalato come l'uso di sistemi di contenimento per impedire le cadute dal letto finisca per creare quella riduzione della motilità globale con decadimento delle condizioni psico-fisiche che è la premessa per le cadute.

Una prima, semplice ma importantissima attività di prevenzione consiste nella rimozione dei numerosi fattori di rischio ambientali, spesso domestici, di cui si ha poca conoscenza. L'attenzione, come abbiamo visto, va concentrata soprattutto sulle scale, sui tappeti, sui cavi volanti, sulle pavimentazioni, sulle soglie non livellate. Deve essere categoricamente sconsigliato il cammino su superfici sdrucciolevoli (per l'uso di cera o su pavimenti bagnati) e l'impiego di vestiti da camera troppo lunghi e di pantofole con suola non rigida. Un secondo livello di intervento dovrà essere di carattere sanitario e consistere nella precoce individuazione e correzione di quelle patologie, talora quasi asintomatiche, che accentuano il rischio di cadute.

D'altra parte va valutato attentamente nell'anziano l'uso di alcuni farmaci che, come abbiamo visto, possono determinare transitorie diminuzioni di pressione arteriosa cerebrale e quindi cadute. L'effettiva indispensabilità di ogni farmaco deve essere accertata e i vantaggi della sua somministrazione devono essere confrontati con l'eventuale tendenza a provocare cadute. Nel caso di prescrizione di psicofarmaci e ipotensivi il paziente deve essere avvertito dell'aumentato rischio di cadute. Tra gli ipnotici è importante dare la preferenza a quelli a breve durata d'azione che tendono ad assicurare solo il riposo notturno.

Lo scadimento senile della qualità del tessuto osseo può essere prevenuto con accorgimenti su cui c'è oggi un generale consenso (attività fisica e dieta appropriata) e forse, in casi selezionati e a rischio, con interventi farmacologici per cui tuttavia non esiste a tutt'oggi un protocollo uniformemente accettato.

L'esercizio fisico costante, inteso sia come deambulazione ma anche come programmi di veri e propri esercizi ginnici e/o nuoto, è un riconosciuto e sicuro fattore di mantenimento della qualità ossea mediato probabilmente attraverso il carico e le forze muscolari.

Nella dieta l'assunzione di calcio ha un ruolo fondamentale. L'apporto quotidiano dovrebbe essere di 1 grammo nelle donne prima della menopausa e di 1,5 grammi in quelle che hanno superato la menopausa. Tale dose può facilmente non essere garantita, per esempio, in chi fa un consumo limitato di latte e dei suoi derivati. L'alcool rappresenta invece, se assunto in dosi elevate e croniche, un sicuro elemento dietetico negativo per la qualità ossea.

A proposito di eventuali interventi farmacologici, infine, il primo problema è rappresentato dalla individuazione dei pazienti a rischio per i quali il rapporto costo/beneficio del trattamento sia vantaggioso. In questo senso un significativo progresso è stato compiuto con le tecniche di misurazione oggettiva del contenuto minerale dell'osso (mineralometria e densitometria ossea). È stato proposto un valore soglia di contenuto minerale osseo di 0,650 grammi per centimetro cubo, valore al di sotto del quale inizia ad aumentare il rischio di una frattura del collo del femore.

I possibili farmaci a disposizione, oltre al calcio, sono il fluoruro di sodio, la vitamina D, gli ormoni anabolizzanti, gli ormoni estrogenici, la calcitonina, l'ipriflavone e i difosfonati, eventualmente anche in associazione tra loro. Non esiste tuttavia un protocollo generalizzabile e l'uso di questi farmaci viene meglio risolto su base individuale. Alcuni di essi sono infatti gravati da effetti secondari e/o rischi la cui valutazione deve essere fatta caso per caso.

BIBLIOGRAFIA

LEWIS FENTON A., *Fracture of Neck of the Femur: Changing Incidence* in «British Medical Journal», 283, pp. 1217-1220, 1981.

WALLACE W. A., *The Increasing Incidence of Fractures of the Proximal Femur: An Orthopaedic Epidemic* in «The Lancet», pp. 1413-1414, 1983.

ELABDIEN ZAIN B. S., OLERUD SVEN, KARLSTROM GORAN e SMEDBY BJORN, *Rising Incidence of Hip Fracture in Uppsala* in «Acta Orthopaedica Scandinavica», 55, pp. 284-289, 1984.

BOYCE W. J. e VESSEY M. P., *Rising Incidence of Fracture of the Proximal Femur* in «The Lancet», pp. 150-151, 1985.

GUPTA K., *Cadute in età avanzata* in «The Practitioner», edizione italiana, 97, pp. 53-55, 1986.

VETERE CARLO, *Le cadute degli anziani* in «Geriatrics», 5, n. 1, pp. 9-15, 1988.

CANIGGIA MARIO e MORREALE PIETRO, *Epidemiology of Hip Fractures in Siena, Italy, 1975-1985* in «Clinical Orthopaedics and Related Research», 238, pp. 131-138, 1989.

PIETROGRANDE V., ALBISETTI W., PIETROGRANDE L. e TORRI G., *Le fratture dell'estremo prossimale del femore, oggi* in «Giornale italiano di ortopedia e traumatologia», 15, supplemento al fascicolo 3, pp. 31-44, 1989.

Le «dita» di zinco

Queste protuberanze proteiche dei fattori di trascrizione, sconosciute fino a circa 10 anni fa, si «afferrano» a specifici siti del DNA regolando in tal modo l'attività dei geni in parecchie specie, dai lieviti all'uomo

di Daniela Rhodes e Aaron Klug

Uno degli interrogativi più affascinanti della biologia contemporanea è come vengano attivati i geni negli organismi pluricellulari. Perché un gene sia attivato è necessario che parecchie proteine, i fattori di trascrizione, si attacchino a un suo segmento, il promotore. L'insieme fattore di trascrizione-promotore forma una specie di «interruttore» per l'attivazione, cioè permette a un enzima di trascrivere un secondo segmento genetico da DNA in RNA. Nella maggior parte dei casi la molecola di RNA risultante funge da stampo per la sintesi di una proteina specifica; talvolta il prodotto finale è ancora RNA. Da tempo ci si chiede come un fattore di trascrizione riesca a distinguere esattamente, su un promotore, il sito particolare a cui deve ancorarsi, distinguendolo dalla massa di DNA di differente natura presente in una cellula.

Oggi si cominciano a trovare le risposte. Molti fattori di trascrizione presentano piccole protuberanze, perfettamente adatte al riconoscimento del DNA, che prendono il nome di «dita» di zinco. Nel 1985 il nostro gruppo di ricerca al Medical Research Council di Cambridge in Inghilterra identificò per primo una di queste strutture in un fattore di trascrizione di rana. Da allora è stato dimostrato che oltre 200 proteine, molte delle quali sono fattori di trascrizione, incorporano le dita di zinco, e parecchi altri fattori di trascrizione contengono strutture, o motivi, affini. Di recente numerosi gruppi, tra i quali il nostro, hanno cominciato a chiarire come le dita di zinco e le strutture a esse correlate riescano a selezionare specifici siti di legame sul DNA e ad ancorarvisi.

Naturalmente, le dita di zinco non sono le uniche strutture che i fattori di trascrizione utilizzano per interagire con il DNA. Altri esempi importanti di strutture sono i motivi elica-ansa-elica (scoperti prima delle dita di zinco, nel 1981), gli omeodominî e le chiusure lampo a dentini di leucina (si vedano gli articoli *Chiusure lampo molecolari nella regolazione genica* di Steven Lanier McKnight in «Le Scienze» n. 274, giugno 1991, e *Geni intelligenti* di Tim Beardsley in «Le Scienze» n. 278, ottobre 1991). Tuttavia le dita di zinco sono di gran lunga i più comuni fra i motivi che si legano al DNA.

La ricerca nel campo del riconoscimento del DNA dovrebbe far progredire le indagini anche in quel campo più vasto che si occupa dello sviluppo degli organismi pluricellulari. In un embrione tutte le cellule contengono gli stessi geni, ma alcune si differenziano diventando, per esempio, neuroni, mentre altre diventano cellule cutanee. I loro destini divergono perché, via via che l'embrione si sviluppa, differenti combinazioni di geni vengono attivate nelle diverse cellule, il che induce la sintesi di quelle proteine specializzate che conferiscono alle cellule differenziate le loro proprietà specifiche. Per comprendere l'attivazione selettiva dei geni sarebbe di importanza fondamentale sapere come i fattori di trascrizione riconoscano i rispettivi siti di legame sul DNA.

Abbiamo scoperto l'esistenza delle dita di zinco dopo esserci interessati dei risultati che provenivano dai laboratori di Robert G. Roeder, allora alla Washington University, e di Donald D. Brown della Carnegie Institution of Washington a Baltimora. All'inizio degli anni ottanta i gruppi di questi due ricercatori avevano analizzato per la prima volta le fasi che portano alla trascrizione di un gene in un organismo più progredito dei batteri. Essi avevano dimostrato che, nell'anuro *Xenopus laevis*, una proteina denominata fattore di trascrizione IIIA (TFIIIA) è un membro di un gruppo di almeno tre fattori che sono necessari per attivare il gene che dà origine all'RNA 5S. Questo RNA è un costituente dei ribosomi, gli organelli su cui le molecole di RNA messaggero (i prodotti della trascrizione genica) vengono tradotte in proteine.

I gruppi di Roeder e Brown hanno poi trovato che il TFIIIA si lega a un segmento relativamente lungo di DNA, che comprende una particolare sequenza di circa 45 coppie di basi della molecola di DNA. (Il DNA consta di due filamenti di nucleotidi contenenti un residuo di desossiribosio, un gruppo fosfato e un residuo di una delle seguenti basi azotate: adenina, citosina, guanina o timina. I due filamenti sono uniti tra loro mediante le basi: l'adenina si appaia sempre con la timina e la citosina con la guanina.)

La lunghezza del sito a cui si attacca il TFIIIA ci ha sorpreso perché la molecola in sé ha dimensioni piuttosto piccole. I fattori di trascrizione delle stesse dimensioni che erano stati identificati in precedenza nei batteri si ancorano a segmenti di DNA molto più corti, dell'ordine di 15 coppie di basi. Come può, ci siamo chiesti, la piccola molecola del TFIIIA occupare un segmento così esteso di DNA? Per fortuna il problema appariva risolubile. I fattori di trascrizione tendono in genere a essere prodotti in scarsa quantità, ma nelle ovaie di *Xenopus* immaturi il TFIIIA è invece abbondante. Qui esso viene immagazzinato sotto forma di complesso con l'RNA 5S che contribuisce a produrre. La sua abbondanza ci ha dato la certezza di poter raccogliere una quantità del complesso TFIIIA-RNA 5S che fosse sufficiente per isolare la proteina. Fatto questo, avremmo potuto ricavare qualche informazione in più sull'organizzazione tridimensionale del TFIIIA e sul modo in cui questa proteina si lega al sito bersaglio sul gene per l'RNA 5S.

Il nostro piano era valido, ma presto incontrammo una difficoltà che si sarebbe dimostrata provvidenziale perché ci mise direttamente sulla via della scoperta delle dita di zinco. Nel 1982 Jonathan Miller, allora borsista presso il nostro laboratorio, si servì di una tecnica di recupero ben nota per estrarre il complesso TFIIIA-RNA 5S dalle ovaie di *Xenopus*, ma, con sua grande delusione, ne ottenne una minima quantità. Si vide poi che il metodo da lui utilizzato eliminava un metallo necessario per tenere unito il complesso. Dopo aver modificato il procedimento di estrazione ed essersi procurato un buon quantitativo del complesso, Miller riuscì a dimostrare che il metallo mancante era lo zinco. Ogni unità TFIIIA-RNA 5S incorporava dai 7 agli 11 ioni zinco, un numero insolitamente grande. Altri esperimenti ci portarono ancora più vicini alla scoperta delle dita di zinco. Scindendo la molecola del TFIIIA in frammenti successivamente più piccoli per mezzo di un enzima, la proteasi, si è visto che i frammenti differivano per grandezza di circa tre chilodalton; alla fine abbiamo ottenuto unità da tre chilodalton resistenti a ogni ulteriore attacco, presumibilmente perché erano ripiegate in modo compatto. Nel loro complesso, questi risultati indicano che la struttura del TFIIIA è quasi interamente costituita da una successione di segmenti da tre chilodalton disposti in tandem (con circa 30 amminoacidi per segmento), ciascuno dei quali è ripiegato attorno a uno ione zinco a formare un piccolo dominio compatto che si lega al DNA.

Se la nostra interpretazione era corretta, ciò significava che ci eravamo imbattuti in un nuovo tipo di fattore di trascrizione. Si era trovato che tutti gli altri fattori studiati in analogo dettaglio interagivano con il DNA come dimeri in cui ciascuna proteina costituente entrava in contatto con il DNA tramite un unico motivo di legame. I nostri risultati sottintendevano che il TFIIIA si allungasse sulla doppia elica, toccandola in parec-

Tre «dita» di zinco che si protendono da un fattore di trascrizione (proteina deputata alla regolazione genica) si fissano all'ampio solco principale di una molecola di DNA. Queste strutture uniscono i fattori di trascrizione ai rispettivi geni bersaglio legandosi a sequenze specifiche di coppie di basi del DNA. Le dita di zinco sono così chiamate perché possono «afferrare» il DNA e perché lo ione zinco al loro centro svolge una funzione decisiva nel determinarne la struttura.

chi punti e non soltanto in uno o due. Questi contatti multipli spiegherebbero anche come il TFIIIA possa interagire con un segmento molto lungo di DNA.

Mentre stavamo cercando conferme al nostro modello, il gruppo di Roeder pubblicò la sequenza amminoacidica del TFIIIA, nella quale trovammo un sostegno alla nostra ipotesi: i primi tre quarti della proteina formavano una serie continua di nove unità simili, costituite da circa 30 amminoacidi. Inoltre, all'interno di ciascuna di queste unità, una coppia di residui di cisteina e una coppia di residui di istidina erano situate in posizioni pressoché identiche. Quest'ultimo risultato concordava con l'idea che ciascuna unità contenesse il proprio ione zinco, in quanto nelle proteine lo zinco si trova generalmente legato a quattro amminoacidi, spesso quattro residui di cisteina o una combinazione di residui di cisteina e di istidina.

Fin dal 1985 questi risultati spinsero uno di noi (Klug) a proporre formalmente che i residui invarianti di cisteina e di istidina fossero utilizzati per ripiegare in maniera indipendente ciascuna unità in un minidominio di legame con il DNA, dominio chiamato in seguito «dito di zinco» perché serviva ad afferrare la doppia elica del DNA. Klug propose anche che la coppia di residui di cisteina vicino a un estremo dell'unità e la coppia di residui di istidina vicino all'altro estremo si legassero allo stesso ione zinco, costringendo la sequenza amminoacidica interposta a formare un'ansa. Così, in ogni unità di 30 amminoacidi, circa 25 amminoacidi si ripiegherebbero in un dominio strutturato (un dito), mentre i restanti amminoacidi servirebbero come collegamento tra dita consecutive.

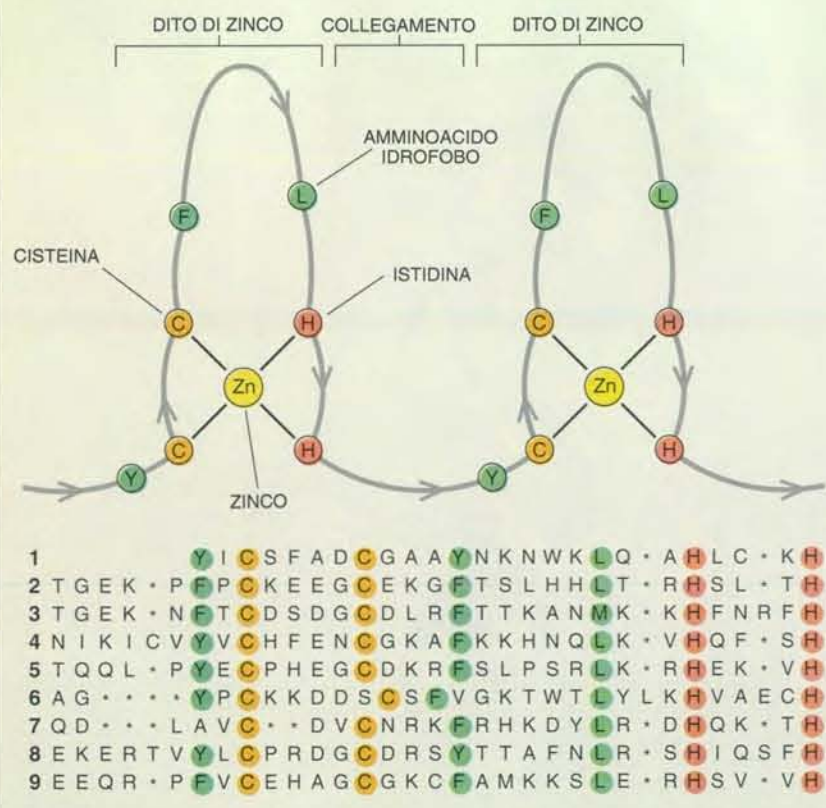
Poco tempo dopo, misurazioni effettuate da noi in collaborazione con Gregory P. Diakun del Daresbury Laboratory del Science and Engineering Research Council vicino a Manchester, in Inghilterra, hanno dimostrato che in effetti ciascuna delle nove unità conteneva uno ione zinco legato a due residui di cisteina e a due residui di istidina. Pertanto la struttura molecolare del TFIIIA era quasi interamente determinata dalla presenza delle nove dita di zinco consecutive. Tutte avevano la stessa architettura di base, ma si distinguevano per la composizione chimica: infatti erano diversi gli amminoacidi che non partecipavano direttamente alla costruzione della struttura del dito.

Ma le dita di zinco entrano effettivamente in contatto con il DNA in maniera indipendente l'una dall'altra, come era stato previsto? Per trovare una risposta, Louise Fairall del nostro gruppo, assieme a ricercatori di altri laboratori, ha effettuato studi di «rilevamento di impronte». Una proteina viene fatta ancorare al DNA; si aggiungono quindi enzimi o altri agenti che attaccano il DNA e lo scindono. Si può presumere che qualunque sito che resista alla scissione sia protetto dalla proteina legata, il che indica che l'area risparmiata è un sito di interazione tra la proteina e il DNA. Nel 1986 i dati ottenuti con questo metodo confermarono l'esistenza di punti di contatto multipli tra TFIIIA e DNA.

Il TFIIIA era, dunque, un'autentica novità, come avevamo sospettato: si connetteva a una regione specifica del DNA utilizzando una serie di moduli di legame indipendenti. L'economia della configurazione modulare era stupenda. Si sapeva già che le cellule si procurano un vasto repertorio di interruttori per attivare i geni combinando variamente un insieme limitato di fattori di trascrizione. Ciò significa che un gene potrebbe essere attivato da una combinazione di proteine a, b e c, mentre un altro gene potrebbe utilizzare soltanto a e b, oppure a, b e d. Con una simile strategia un organismo evita di dover produrre uno specifico fattore di trascrizione per ciascuno dei numerosissimi geni che sono attivi nelle cellule. Gli studi sulle dita di zinco hanno rivelato che il principio combinatorio funziona anche all'interno di un fattore di trascrizione. Una cellula può produrre una serie molto vasta di fattori di trascrizione ben distinti variando la scelta, l'ordine e il numero di moduli indipendenti in grado di legarsi al DNA. Una particolare combinazione di dita di zinco in un fattore di trascrizione gli consente di riconoscere una e una sola sequenza specifica sul DNA.

La scoperta delle dita di zinco

Nel 1985 uno degli autori (Klug) scoprì che certi segmenti di sequenze amminoacidiche possono avvolgersi in maniera indipendente attorno a uno ione zinco, formando strutture che hanno successivamente preso il nome di dita di zinco. Le curve in grigio rappresentano un filamento amminoacidico; i cerchietti colorati indicano invece quegli amminoacidi che Klug correttamente riteneva potessero partecipare all'avvolgimento. Un indizio importante sul tipo di avvolgimento emerse dall'esame della sequenza degli amminoacidi (*lettere maiuscole in basso*) nel TFIIIA. La maggior parte di questa proteina è costituita da nove unità successive (*numerate*), che mostrano somiglianze importanti: includono, in posizioni pressoché identiche, una coppia di residui di cisteina (*C*), una coppia di residui di istidina (*H*) e, con la possibile eccezione della unità 7, tre residui di amminoacidi idrofobi (*in verde*). (Gli asterischi indicano interruzioni prive di importanza.) Queste osservazioni, assieme a dati biochimici, hanno permesso di formulare l'ipotesi che le coppie di residui di cisteina e di istidina presenti in ogni motivo si combinino con un singolo ione zinco (Zn), provocando così la formazione di un'ansa nel filamento amminoacidico. Nel contempo, i tre amminoacidi idrofobi contribuirebbero in qualche modo a stabilizzare l'insieme.

La struttura di un dito di zinco (*a sinistra*) è stata ricostruita nei dettagli alla fine degli anni ottanta. Il «nastro» rappresenta lo scheletro carbonio-azoto della catena di amminoacidi. La metà sinistra si ripiega all'indietro su se stessa per formare una sottostruttura a due filamenti, il foglietto beta, mentre la metà destra si avvolge in un'alfa elica. Il legame con lo zinco (*sfera gialla*) dei residui di cisteina nel foglietto beta (*in giallo*) e dei residui di istidina nell'elica (*in rosso*) riunisce le due metà vicino alla base del dito; avvicina l'uno all'altro anche gli amminoacidi idrofobi (*in verde*) posti alla sommità del dito, dove la loro reciproca attrazione contribuisce a mantenere intatta la struttura. A destra, le tre dita di zinco in tandem (*nastro rosso con strisce bianche*) di Zif268, una proteina deputata alla regolazione genica, hanno tutte stabilito un contatto (*in viola*) con le basi che si trovano nel solco principale del DNA (*in blu*), ancorandosi collettivamente a quasi un intero avvolgimento della doppia elica. In questa rappresentazione sono visibili cinque dei sei contatti presi con le basi. Le linee e gli anelli in giallo indicano connessioni tra lo zinco e i residui di cisteina e di istidina. L'immagine di Zif268 si basa su un'analisi cristallografica a raggi X effettuata da Nikola P. Pavletich e Carl O. Pabo della Johns Hopkins University.

L'efficienza del metodo combinatorio ci ha indotto a pensare che il dito di zinco possa essere un motivo presente in molte proteine; di fatto, negli eucarioti è sorprendentemente ubiquitario. Peter F. R. Little dell'Imperial College di Londra stima che, nelle cellule umane, fino all'1 per cento del DNA codifichi per le dita di zinco; nel cromosoma 19 questa cifra sale all'8 per cento. Le proteine contenenti dita di zinco finora identificate presentano da due fino a 37 di queste strutture disposte in tandem.

Per capire come un dito di zinco riconosca una sequenza specifica di coppie di basi (che adotta una conformazione specifica), occorre conoscere nei particolari la sua struttura tridimensionale. La maggior parte delle proteine include regioni locali con struttura «secondaria» che, avvolgendosi, danno luogo alla conformazione tridimensionale globale della proteina. Le strutture secondarie più comuni sono l'alfa elica (in cui lo scheletro della proteina si avvolge formando un'elica caratteristica) e il filamento beta (in cui l'impalcatura di sostegno è completamente distesa) (si veda l'articolo *Il problema dell'avvolgimento delle proteine* di Frederic M. Richards in «Le Scienze» n. 271, marzo 1991).

Nel 1988 Jeremy M. Berg della Johns Hopkins University chiarì su basi teoriche gli aspetti principali della architettura tridimensionale delle dita di zinco, ma il suo modello non venne confermato fino all'anno successivo. Peter E. Wright e collaboratori della Scripps Clinic and Research Foundation di La Jolla, in California, determinarono poi la struttura di un dito di zinco della proteina Xfin di *Xenopus laevis* mediante spettroscopia di risonanza magnetica nucleare (NMR), una tecnica che può essere utilizzata per risolvere la struttura tridimensionale di piccole proteine in soluzione. Poco dopo altri gruppi di ricerca, compreso il nostro, identificarono la stessa struttura in altre proteine con dita di zinco.

Come aveva previsto Berg, la sequenza amminoacidica caratteristica di un dito di zinco si ripiega in una conformazione compatta, dando origine, nello stesso tempo, a due sottostrutture ben evidenti. Una parte della sequenza (comprendente, per esempio, la metà sinistra di una prominenza verticale) adotta la forma di un «foglietto» beta: si compone di due filamenti beta che danno origine al foglietto quando il secondo filamento si ripiega sul primo. L'altra parte della sequenza (la metà «destra») si avvolge a elica, dando origine a un'alfa elica. I due residui di cisteina si trovano nella parte inferiore del foglietto beta, mentre i due residui di istidina si trovano al fondo dell'elica. Tutti e quattro sono uniti per mezzo di un atomo di zinco, che serve essenzialmente a fissare assieme il foglietto beta e l'elica.

La spettroscopia di risonanza magnetica nucleare ha anche contribuito a chiarire il ruolo di alcuni altri amminoacidi. Quando in origine abbiamo esaminato la sequenza del TFIIIA, abbiamo notato che le presunte dita includevano tutte un gruppo di tre amminoacidi idrofobi in posizioni pressoché identiche. (Le sostanze idrofobe si associano spesso l'una con l'altra nell'interno di una proteina, lontano dall'ambiente acquoso circostante.) Questa invarianza ha fatto supporre che i suddetti amminoacidi avessero un importante ruolo strutturale. Benché nelle rappresentazioni lineari della sequenza amminoacidica appaiano piuttosto distanti l'uno dall'altro, abbiamo pensato che potessero in qualche modo interagire nello spazio tridimensionale e partecipare all'avvolgimento del minidominio. Coerentemente con il modello di Berg, i risultati della spettro-

scopia di risonanza magnetica nucleare hanno dimostrato che, in effetti, quando il modulo che costituisce il dito di zinco si ripiega, gli amminoacidi idrofobi si avvicinano l'uno all'altro in misura sufficiente a essere soggetti ad attrazione reciproca. Formano così un nucleo idrofobo che contribuisce a stabilizzare la forma del modulo.

Parallelamente ai tentativi fatti per capire la struttura delle dita di zinco abbiamo preso in considerazione, assieme ad altri ricercatori, un problema più generale. Molti esperimenti ci avevano fatto concludere che nel TFIIIA le dita di zinco, che rappresentano la parte preponderante della molecola, erano da sole responsabili della capacità di quest'ultima di riconoscere il promotore del gene per l'RNA 5S. A poco a poco, però, si è trovato un numero sempre crescente di proteine nella cui grossa molecola erano presenti solo poche dita di zinco. Come potevano queste brevi serie di dita di zinco dirigere la proteina verso un promotore senza la cooperazione di altre parti della molecola?

Nei nostri sforzi per dare una risposta a questi interrogativi abbiamo concentrato l'attenzione su un fattore di trascrizione del lievito dotato di tre dita di zinco: il fattore SWITCH 5 (SWI5). Con il collega Kyoshi Nagai, abbiamo isolato la regione delle dita e l'abbiamo messa in presenza del promotore del gene bersaglio di SWI5. Com'era prevedibile, il segmento isolato si è unito prontamente al promotore, il che sottintende che le dita di zinco siano da sole responsabili del legame con il DNA. Un fatto interessante è che, perché la proteina SWI5 possa ancorarsi con forza sufficiente all'esatto sito bersaglio sul DNA, devono essere presenti almeno due dita di zinco collegate. Esaminando con la spettroscopia di risonanza magnetica nucleare le prime due dita di zinco di SWI5, abbiamo confermato, con i colleghi David Neuhaus e Yukinobu Nakeseko, che dita di zinco adiacenti non si fondono insieme; esse sono proprio «testine di lettura» indipendenti, unite da collegamenti flessibili.

Restavano però da identificare gli esatti punti di contatto tra dita di zinco e DNA. Nikola P. Pavletich e Carl O. Pabo, che allora lavoravano alla Johns Hopkins University, ottennero i primi risultati importanti nel 1991. Per iniziare, prepararono cristalli del complesso formato dal DNA e dal dominio di legame del fattore di trascrizione Zif268; effettuando quindi un'analisi cristallografica a raggi X, riuscirono a determinare nei particolari la struttura di questo complesso. Lo Zif268, che, analogamente a SWI5, include un gruppo di tre dita di zinco, prende parte alle prime fasi di sviluppo degli embrioni di topo.

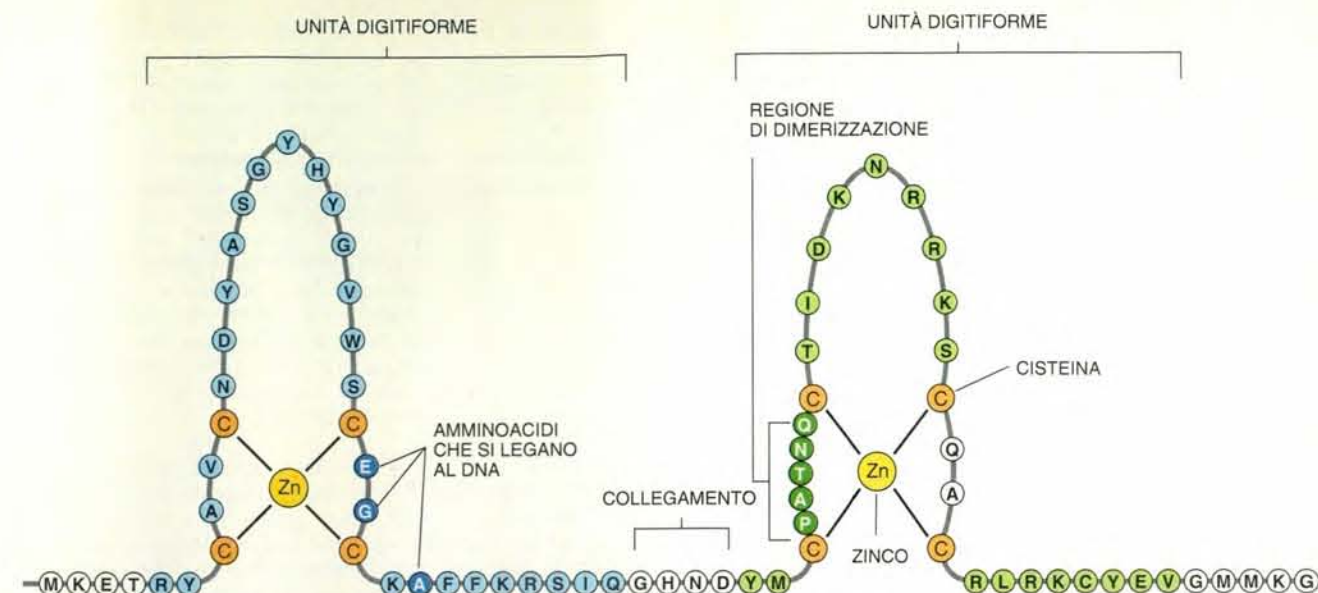
Le analisi ai raggi X hanno rivelato che la regione di Zif268 che contiene le dita di zinco si avvolge per quasi un giro completo attorno all'elica del DNA (più o meno come una lettera C), inserendosi nel solco principale. (Quest'ultimo è la più ampia di due infossature parallele che decorrono a spirale attorno all'asse longitudinale della doppia elica di DNA.) Le dita di zinco stabiliscono contatti con siti successivi, costituiti da tre coppie di basi, del DNA e si accostano tutte alla doppia elica più o meno con la stessa orientazione: l'alfa elica di ogni dito si spinge all'interno del solco principale appoggiandosi a una delle pareti.

Più specificamente, il primo e il terzo dito di Zif268 si legano al DNA in modo identico: un amminoacido nel primo avvolgimento dell'alfa elica entra in contatto con la prima coppia di basi del corrispondente sito di legame sul DNA, e un amminoacido che si trova nel terzo avvolgimento dell'elica entra in contatto con la terza coppia di basi dello stesso sito di legame. Anche il secondo dito stabilisce contatti mediante l'alfa elica, ma questa volta gli amminoacidi che si trovano sul primo e sul secondo avvolgimento entrano in contatto con la prima e con la seconda coppia di basi del corrispondente sito di legame sul DNA. (In ogni caso, un amminoacido entra in contatto con una sola base di una coppia.) Inoltre sia l'alfa elica sia il foglietto beta delle dita di zinco si legano ai gruppi fosfato delle catene fosfoglucidiche che, se si immagina la molecola di DNA come una scala a pioli, ne costituiscono i «montanti». Questi legami supplementari contribuiscono a stabilizzare l'aggancio delle dita di zinco al DNA.

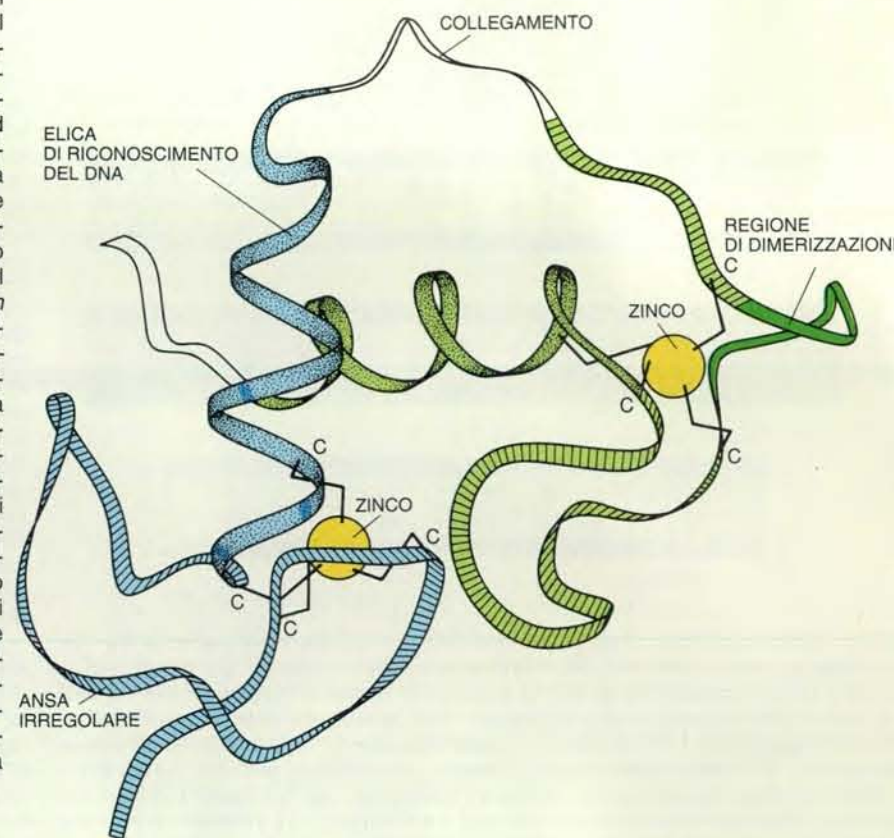
Finora nessun altro complesso formato dalle dita di zinco e dal DNA è stato risolto mediante cristallografia a raggi X. Nondimeno, Grant H. Jacobs del nostro laboratorio ha raccolto prove convincenti del fatto che molte dita di zinco possono legarsi al DNA più o meno nello stesso modo di Zif268. Egli ha confrontato le sequenze amminoacidiche di oltre 1000 dita di zinco e ha trovato che gli amminoacidi in tre posizioni sono particolarmente variabili. Queste posizioni a elevata variabilità sono esattamente quelle che vengono utilizzate per stabilire i contatti nel complesso di Zif268, cioè quelle che si situano sul primo, il secondo e il terzo avvolgimento dell'alfa elica. Una tale somiglianza prospetta l'affascinante eventualità che un giorno si possano progettare a piacere dita di zinco in grado di riconoscere particolari sequenze di DNA, un risultato che potrebbe avere importanti risvolti teorici e pratici.

Naturalmente vi sono limiti a ciò che si può estrapolare dal modello dello Zif268 e dalle analisi statistiche. Si può pensare che le proteine con molte dita di zinco interagiscano con il DNA in maniera un poco diversa. Per esempio, se il tipo di legame dello Zif268 fosse valido anche per il TFIIIA, questa proteina, con le sue nove dita di zinco, si avvolgerebbe attorno al DNA per tre giri, come il filo su una spoletta. La stessa lunghezza di tale avvolgimento potrebbe impedire al fattore di staccarsi facilmente dal DNA quando ciò divenisse necessario. Di fatto, i dati delle analisi di rilevamento delle impronte effettuate da noi e da altri ricercatori indicano che il TFIIIA non si attorciglia in maniera continua attorno al DNA. Le prime tre dita del TFIIIA afferrano quasi certamente un singolo avvolgimento dell'elica di DNA ed è molto probabile che le ultime tre dita facciano lo stesso. Ma la maggior parte della molecola di TFIIIA si situa su un unico lato della doppia elica; pertanto deve attraversare almeno due volte il solco minore. I diversi modi con cui regioni separate del TFIIIA si legano al DNA riflettono probabilmente il fatto che le sequenze amminoacidiche delle dita di zinco di questa proteina differiscono l'una dall'altra più di quelle delle dita presenti in proteine simili a Zif268.

Da un punto di vista evolutivo vi sono buoni motivi per pensare che i dominî di legame con molte dita di zinco abbiano avuto origine per duplicazione di qualche gene ancestrale che specificava una proteina di piccole dimensioni, costituita da una trentina di amminoacidi. Pensiamo anche che questa catena di 30 amminoacidi possa essere stata tra le primissime proteine ad evolversi; in fin dei conti, una molecola del genere sarebbe stata semplice da produrre. Una volta sintetizzata, essa avrebbe catturato senza difficoltà e senza rischi zinco (un metallo relativamente inerte) dall'ambiente circostante, ripiegandosi senza alcun intervento esterno e assumendo così una conformazione stabile. Una volta ripiegata, avrebbe acquisito la capacità di legarsi al DNA o all'RNA. Queste caratteristiche contribuiscono quasi certamente a spiegare perché le dita di zinco siano oggi estremamente diffuse nei regni animale e vegetale. Qualunque specie che acquisisca il programma genetico per sintetizzare un particolare tipo di dita di zinco capace di ripiegarsi in maniera autonoma acquisirebbe istantaneamente anche la capacità di legarsi a un nuovo segmento di DNA. Questa proprietà, a sua volta, potrebbe dare origine a nuove funzioni cellulari, come la capacità di trascrivere qualche gene in precedenza silente e quindi di produrre un nuovo enzima o proteina utile.

Mentre con altri ricercatori stavamo acquisendo una conoscenza sempre più approfondita della struttura e della funzione delle dita di zinco classiche, altri importanti risultati hanno cominciato a indicarci che il motivo scoperto inizialmente nel TFIIIA non era l'unica struttura deputata al riconoscimento del DNA che si basasse sullo zinco. Già nel 1987 alcuni ricercatori avevano determinato le sequenze amminoacidiche di parecchi membri di una grande famiglia di fattori di trascrizione, i recettori nucleari degli ormoni. Per poter attivare un gene, questi fattori devono

Il dominio di legame con il DNA del recettore per gli estrogeni (un fattore di trascrizione che deve essere legato a un estrogeno per agire su un gene) è costituito da una sequenza di amminoacidi che forma due unità simili a dita di zinco (*in blu e in verde*). Inizialmente si supponeva che, come le dita di zinco classiche, ciascuna delle due unità riconoscesse una distinta sequenza di basi del DNA. Ora si ritiene che solo i tre amminoacidi in blu scuro interagiscano con le basi del DNA, il che implica che a stabilire il contatto principale con il DNA sia la prima unità digitiforme. La seconda svolge una diversa funzione: contiene cinque amminoacidi (*in verde scuro*) che permettono a una molecola di recettore di formare un dimero con un altro recettore. L'appaiamento è necessario perché i recettori per gli estrogeni si uniscano saldamente al DNA.

Un modello del dominio di legame con il DNA del recettore per gli estrogeni

La struttura particolareggiata del dominio di legame con il DNA del recettore per gli estrogeni è stata determinata nel 1990 da uno degli autori (Rhodes) e dai suoi collaboratori, John W. R. Schwabe e David Neuhaus. Questi ricercatori sono riusciti a individuare quali parti della struttura tridimensionale svolgano le fondamentali funzioni di riconoscimento del DNA e di appaiamento con altri recettori. Le due unità del dominio di legame (*in blu chiaro e in verde chiaro*) adottano conformazioni simili. Un'ansa irregolare (*tratteggiata*), che include due residui di cisteina (*C*), è seguita da un'alfa elica (*spirale punteggiata*), la quale contiene il terzo e il quarto residuo di cisteina. Il legame dello zinco con i residui di cisteina unisce i segmenti terminali dell'ansa irregolare alla base dell'elica; così avvolte, le due unità si concatenano mediante le loro eliche. Gli amminoacidi responsabili del riconoscimento di basi specifiche (*in blu scuro*) sono localizzati nell'alfa elica della prima unità; quelli responsabili della formazione di un dimero (*in verde scuro*) sono localizzati invece nell'ansa irregolare della seconda unità.

prima unirsi a un particolare ormone di natura steroidea o tiroidea o a una particolare vitamina. Nell'esaminare le sequenze individuate, si è visto che ciascuna conteneva un dominio di circa 80 amminoacidi, identico o estremamente simile in tutte le molecole. Questo dominio includeva costantemente due, e sempre due, unità la cui sequenza amminoacidica ricordava quella delle dita di zinco. Come per queste ultime, ogni unità, o motivo, conteneva due coppie di amminoacidi potenzialmente in grado di legarsi allo zinco; nel caso in esame, tuttavia, si trattava esclusivamente di residui di cisteina, e non di cisteina e istidina. Le somiglianze di queste sequenze con le dita di zinco del TFIIIA implicavano che il segmento da 80 amminoacidi ricco di cisteina, che si trova nei fattori di trascrizione, fosse il dominio che si lega al DNA.

Pierre Chambon e Stephen Green dell'INSERM di Strasburgo confermarono quest'ipotesi alla fine degli anni ottanta, e poco dopo Paul B. Sigler, allora all'Università di Chicago, Keith R. Yamamoto dell'Università della California a San Francisco e i loro collaboratori stabilirono che ciascuno dei due segmenti del dominio che si lega al DNA incorpora un atomo di zinco. Naturalmente noi e altri ci aspettavamo che, come accade per le dita di zinco simili a quelle del TFIIIA, le configurazioni dei due motivi si assomigliassero ed essi formassero moduli in grado di legarsi al DNA in modo indipendente.

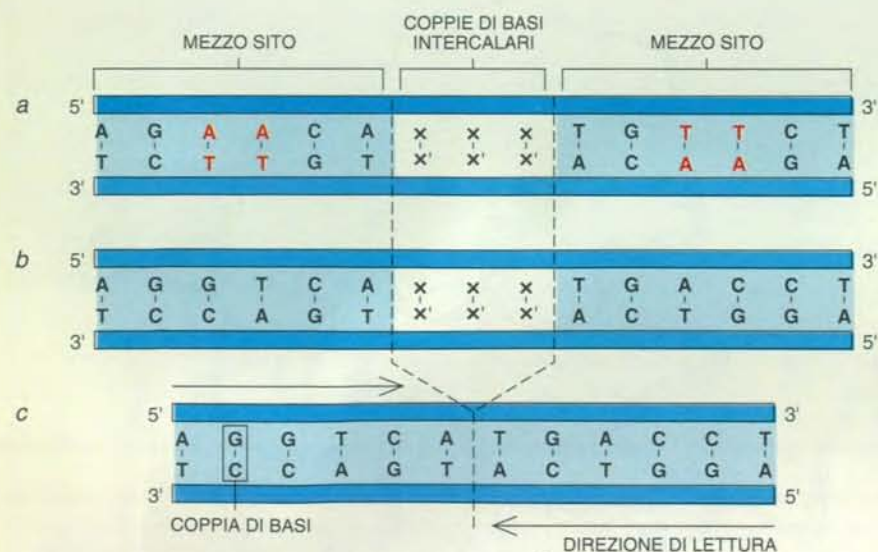
L'ipotesi è risultata in parte errata. Le analisi strutturali hanno infine indicato che le due unità si ripiegano in modo simile, ma in precedenza alcuni sorprendenti risultati avevano dimostrato che esse non funzionano come testine di lettura indipendenti del DNA. Sostituendo un amminoacido con un altro ed esaminando l'effetto di questa sostituzione sul legame con il DNA, Chambon, Ronald M. Evans del Salk Institute for Biological Studies di San Diego e Gordon M. Ringold, in precedenza della Stanford University, hanno trovato, assieme ai loro collaboratori, che il primo motivo funge da unità principale di riconoscimento del DNA. All'incirca nello stesso periodo, Evans e il suo collaboratore Kazuhiko Umesono, ricorrendo di nuovo al metodo di sostituzione, hanno scoperto almeno una delle funzioni del secondo motivo. Per comprenderla, si devono innanzitutto acquisire alcune nozioni generali sul modo in cui i recettori per gli ormoni steroidei interagiscono con il DNA.

Questi recettori si legano al DNA in coppie, o dimeri, di molecole identiche. Ciascuna proteina di una coppia riconosce metà di un sito di legame che consta di due parti, denominato palindromo in quanto le due metà sono identiche se lette in direzioni opposte, cioè scorrendo lungo i filamenti opposti del DNA. La sequenza di basi del mezzo sito riconosciuto da un tipo di fattore di trascrizione (per esempio il recettore per gli estrogeni) può corrispondere esattamente a quella riconosciuta da un altro fattore (per esempio il recettore per gli ormoni tiroidei). In questo caso l'unica differenza tra i due siti di legame è il numero di coppie di basi che separano le due metà di ogni sito palindromo.

Di conseguenza un fattore di trascrizione, per poter trovare sul DNA il proprio sito di ancoraggio, deve contenere regioni deputate sia a distinguere la sequenza di basi specifica di una metà di un sito sia a misurare la distanza tra le due metà del sito stesso. Evans e Umesono hanno trovato che una parte del secondo motivo serve proprio a misurare questa spaziatura.

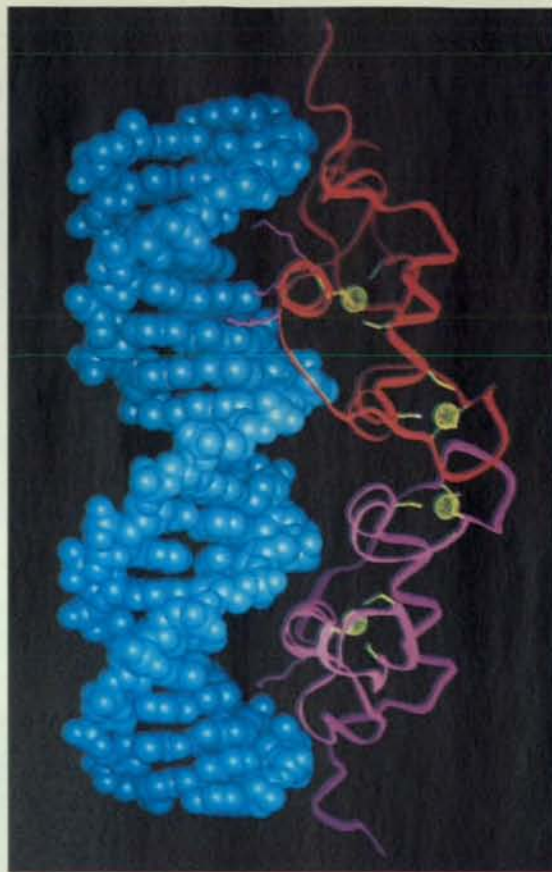
Malgrado i progressi compiuti, non si era ancora in grado di riunire in un tutto unico queste informazioni sparse, per spiegare in che modo avvenisse il riconoscimento specifico di una sequenza. A questo scopo era indispensabile conoscere la configurazione tridimensionale dei domini di legame con il DNA dei recettori proteici, per poter vedere in quali punti della struttura si collocassero gli amminoacidi responsabili di questa funzione. Nel 1990, mediante la spettroscopia di risonanza magnetica nucleare, Robert Kaptein e collaboratori dell'Università di Utrecht, nei Paesi Bassi, hanno risolto la struttura del dominio di legame di un recettore per i glucocorticoidi (cortisone) presente nel ratto. Poco dopo John W. R. Schwabe e Neuhaus hanno risolto, assieme a uno di noi (Rhodes), la struttura del dominio di legame del recettore umano per gli estrogeni.

Come si poteva dedurre dalla composizione amminoacidica simile, si è visto che i domini di legame adottano la stessa struttura sia nei recettori per i glucocorticoidi sia in quelli per gli estrogeni. Ciascuno dei due motivi che contengono dita di zinco all'interno del dominio di legame consiste in due parti: un filamento amminoacidico avvolto in maniera irregolare (anziché un foglietto beta come nelle dita di zinco classiche) a cui fa seguito un'alfa elica. La parte avvolta contiene due siti di legame per lo zinco, mentre gli altri due siti si trovano all'inizio della successiva elica. Tuttavia, invece di rimanere separati come fanno le comuni dita di zinco, i due motivi si fondono in una singola unità strutturale. In questa disposizione le eliche si incrociano perpendicolarmente in corrispondenza del loro punto medio, creando una configurazione che è dovuta all'attrazione reciproca degli amminoacidi idrofobi invarianti e relativamente invarianti.

Avendo accertato la struttura tridimensionale del dominio di legame con il DNA, abbiamo continuato il nostro lavoro localizzando su di esso la posizione degli amminoacidi che erano stati identificati come importanti per il riconoscimento del DNA. Gruppi diretti da Chambon, Evans e Ringold avevano identificato nel primo motivo digitiforme tre amminoacidi che erano responsa-

I siti di legame sul DNA, detti anche elementi di risposta, sono costituiti da due mezzi siti che sono riconosciuti dai recettori per i glucocorticoidi (*a*), per gli estrogeni (*b*) e per gli ormoni tiroidei (*c*). In qualunque elemento i mezzi siti sono uguali se le loro coppie di basi (*lettere unite da un trattino*) vengono lette scorrendo lungo i filamenti opposti del DNA (*frecce in c*) nella direzione 5' → 3'. I recettori si legano agli elementi di risposta come coppie, o dimeri (un recettore per ogni mezzo sito). Perché il legame avvenga, essi devono poter distinguere sia la sequenza di basi dei mezzi siti, sia la spaziatura esistente tra essi. Le differenze tra elementi di risposta possono essere tenui: l'elemento *b* differisce da *a* solo per due coppie di basi (*lettere in rosso*) e dall'elemento *c* solo per il numero di basi interposte tra i mezzi siti.

bili del riconoscimento della sequenza di basi di un mezzo sito. Si è visto che questi amminoacidi si trovano sullo stesso lato dell'alfa elica, il che ci induce a chiamare questa sottostruttura «elica per il riconoscimento del DNA». Questo ci rivela qualcosa di più anche sulla funzione del secondo motivo: incrociando l'elica per il riconoscimento del DNA del primo motivo, quella del secondo funge da supporto per mantenerla in sede. La separazione di funzioni tra i due motivi fa pensare che il secondo abbia avuto origine dalla duplicazione del primo, ma, una volta comparso, sia stato reclutato per svolgere nuovi compiti.

La costruzione di mappe tridimensionali ha mostrato come il secondo motivo svolga il compito vitale d'individuare la spaziatura tra mezzi siti sul DNA. Evans e Umesono avevano stabilito che gli amminoacidi responsabili di questa funzione si trovano tra i primi due residui di cisteina del secondo motivo: nella configurazione tridimensionale sono localizzati nel segmento ad anse irregolari che precede l'elica, dove restano disponibili in attesa di legare una molecola alla sua partner. Il ricorso a modelli al calcolatore dell'interazione tra DNA e regioni di legame dei recettori per i glucocorticoidi e per gli estrogeni ha quindi permesso, rispettivamente al gruppo di Kaptein e al nostro, di scoprire che l'appaiamento delle proteine tramite la connessione prevista orienterebbe il dimero in modo corretto. Le due eliche di riconoscimento su quest'ultimo sarebbero disposte in maniera tale che la spaziatura tra di esse corrisponda alla spaziatura tra i due mezzi siti appropriati del DNA.

In seguito Sigler e Ben F. Luisi, suo collaboratore alla Yale University, assieme a Yamamoto e a Leonard P. Freedman dell'Università della California a San Francisco, hanno confermato il modello per mezzo della cristallografia a raggi X. Essi hanno anche osservato che ciascuna proteina del dimero stabilisce parecchi contatti con gruppi fosfato situati su entrambi i lati del solco principale. Questi contatti posizionano l'elica di riconoscimento del DNA in modo tale che possa giungere in profondità nel solco e formare legami con le coppie di basi che si trovano nel mezzo sito. Nel complesso, dunque, gli studi sulle dita di zinco dei recettori per gli ormoni indicano che questi motivi, nonostante somiglianze strutturali con le dita di zinco di proteine del tipo TFIIIA, funzionano in maniera più simile a quella dei motivi di riconoscimento del DNA che si trovano in altri fattori di trascrizione, come l'elica-ansa-elica oppure la chiusura lampo a dentini di leucina. In altri termini, ripiegandosi insieme anziché rimanere distinti, sono di aiuto ai recettori per gli ormoni per formare i dimeri che permettono loro di riconoscere i siti di legame specifici sul DNA.

L'aggancio al DNA è stato effettuato dai domini di legame (*in rosso e in viola*) di una coppia di recettori per i glucocorticoidi. Un'alfa elica (*spirale*) di ciascun dominio entra in contatto (*linee sottili in viola*) con le basi che si trovano su un solo lato della doppia elica di DNA. L'immagine si basa su un'analisi cristallografica a raggi X eseguita da Ben F. Luisi e Paul B. Sigler e collaboratori della Yale University.

Quando la conoscenza della struttura di una molecola rivela qualche cosa sul suo funzionamento, può anche servire a gettare luce su una malattia. Nel caso delle dita di zinco, si è trovato che un carcinoma che colpisce i reni, il tumore di Wilm, ha origine da una mutazione genetica che interferisce con il legame stabilito con il DNA dalla regione a dita di zinco di una particolare proteina. Inoltre, alcuni sintomi che appaiono legati a un'insufficiente assunzione di zinco nella dieta - come un ritardo nello sviluppo sessuale - possono essere attribuiti oggi all'incapacità dei recettori per gli estrogeni e per gli androgeni di avvolgersi in modo corretto in assenza di zinco.

È chiaro che le due classi di dita di zinco di cui abbiamo parlato differiscono profondamente sia per la struttura sia per il modo in cui interagiscono con il DNA e non c'è alcun dubbio che una varietà ancora più grande sarà scoperta nell'ampia famiglia delle proteine dotate di questi motivi. Così, si continuano a scoprire sequenze di amminoacidi comprendenti quelli che sembrano motivi di legame con lo zinco, anche se la spaziatura tra le coppie di residui di cisteina o di istidina, oppure il numero delle coppie differisce rispetto a ciò che si riscontra in un normale dito di zinco. Un esempio insolito è la proteina GAL4 del lievito: essa contiene sei residui di cisteina che si avvolgono attorno a due atomi di zinco. Ci aspettiamo anche di scoprire che alcune dita di zinco, o molecole strettamente affini, siano coinvolte in attività diverse dalla trascrizione, come il trasporto o la rielaborazione del DNA o dell'RNA.

BIBLIOGRAFIA

MILLER J., McLACHLAN A. D. e KLUG A., *Repetitive Zinc-Binding Domains in the Protein Transcription Factor IIIA from* Xenopus *Oocytes* in «EMBO Journal», 4, n. 6, 1985.

KLUG AARON e RHODES DANIELA, *Zinc Fingers: A Novel Protein Motif for Nucleic Acid Recognition* in «Trends in Biochemical Sciences», 12, n. 12, dicembre 1987.

PAVLETICH NIKOLA P. e PABO CARL O., *Zinc Finger-DNA Recognition: Crystal Structure of a ZIF268-DNA Complex at 2.1Å* in «Science», 252, pp. 809-817, 10 maggio 1991.

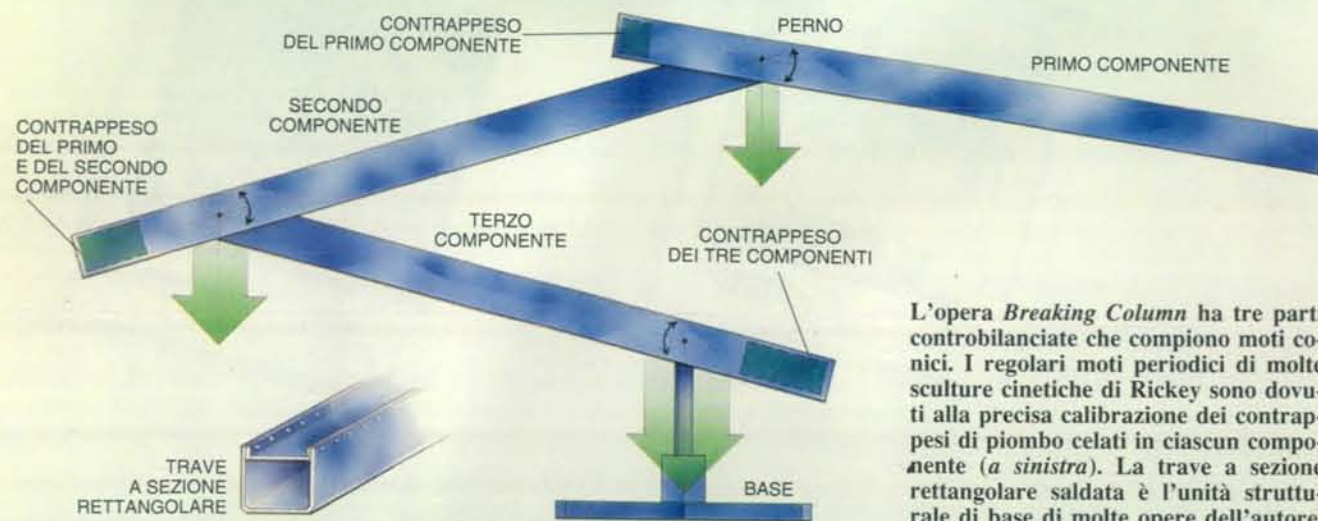
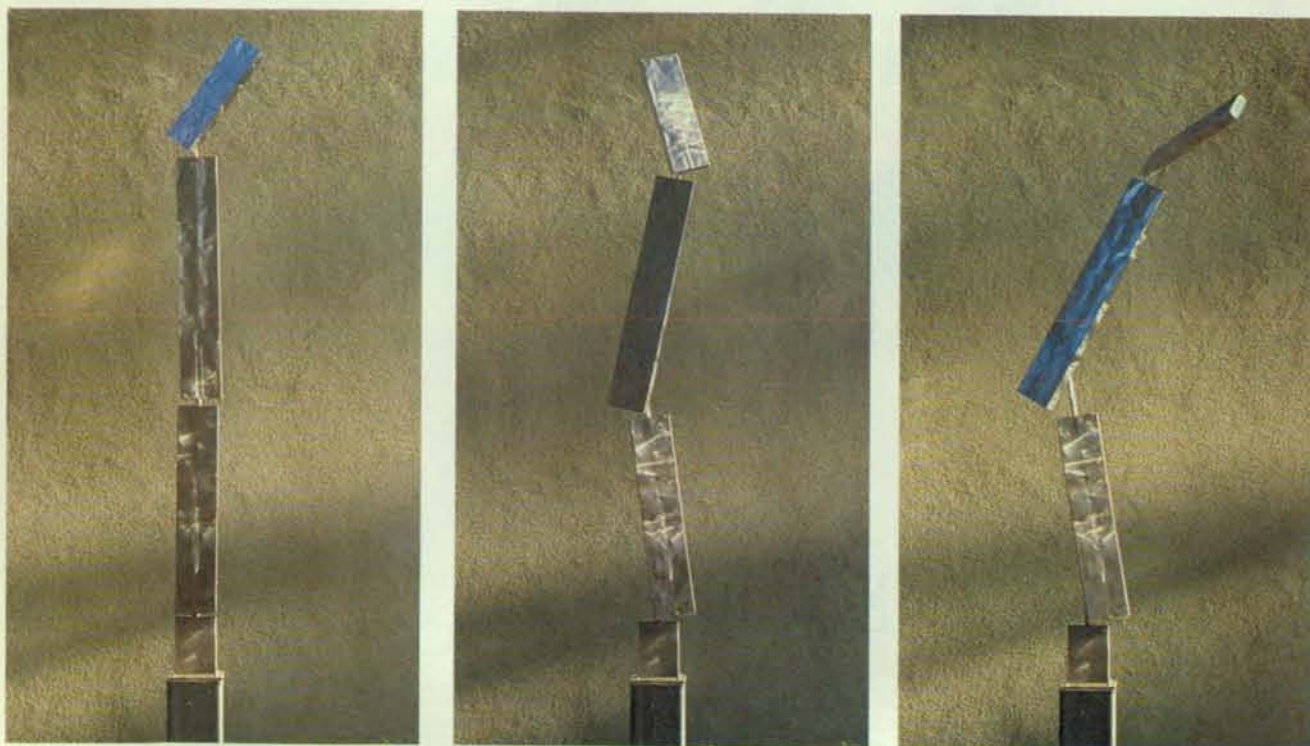
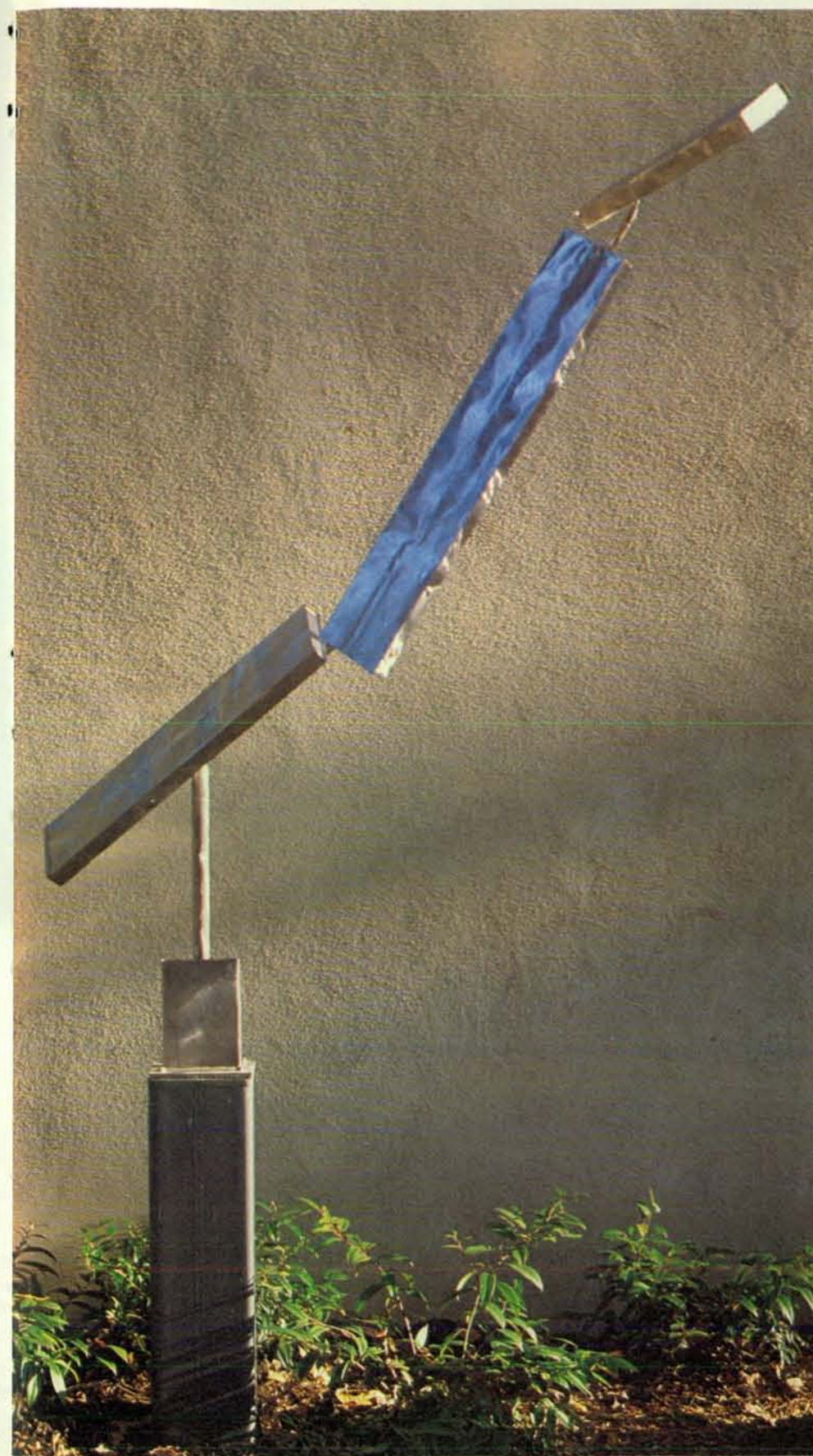
SCHWABE JOHN W. R. e RHODES DANIELA, *Beyond Zinc Fingers: Steroid Hormone Receptors Have a Novel Structural Motif for DNA Recognition* in «Trends in Biochemical Sciences», 16, n. 8, agosto 1991.

LUISI B. F., XU W. X., OTWINOWSKI Z., FREEDMAN L. P., YAMAMOTO K. R. e SIGLER P. B., *Crystallographic Analysis of the Interaction of the Glucocorticoid Receptor with DNA* in «Nature», 352, n. 6335, 8 agosto 1991.

HARRISON STEPHEN C., *A Structural Taxonomy of DNA-Binding Domains* in «Nature», 353, n. 6346, 24 ottobre 1991.

Aspetti tecnologici dell'arte cinetica

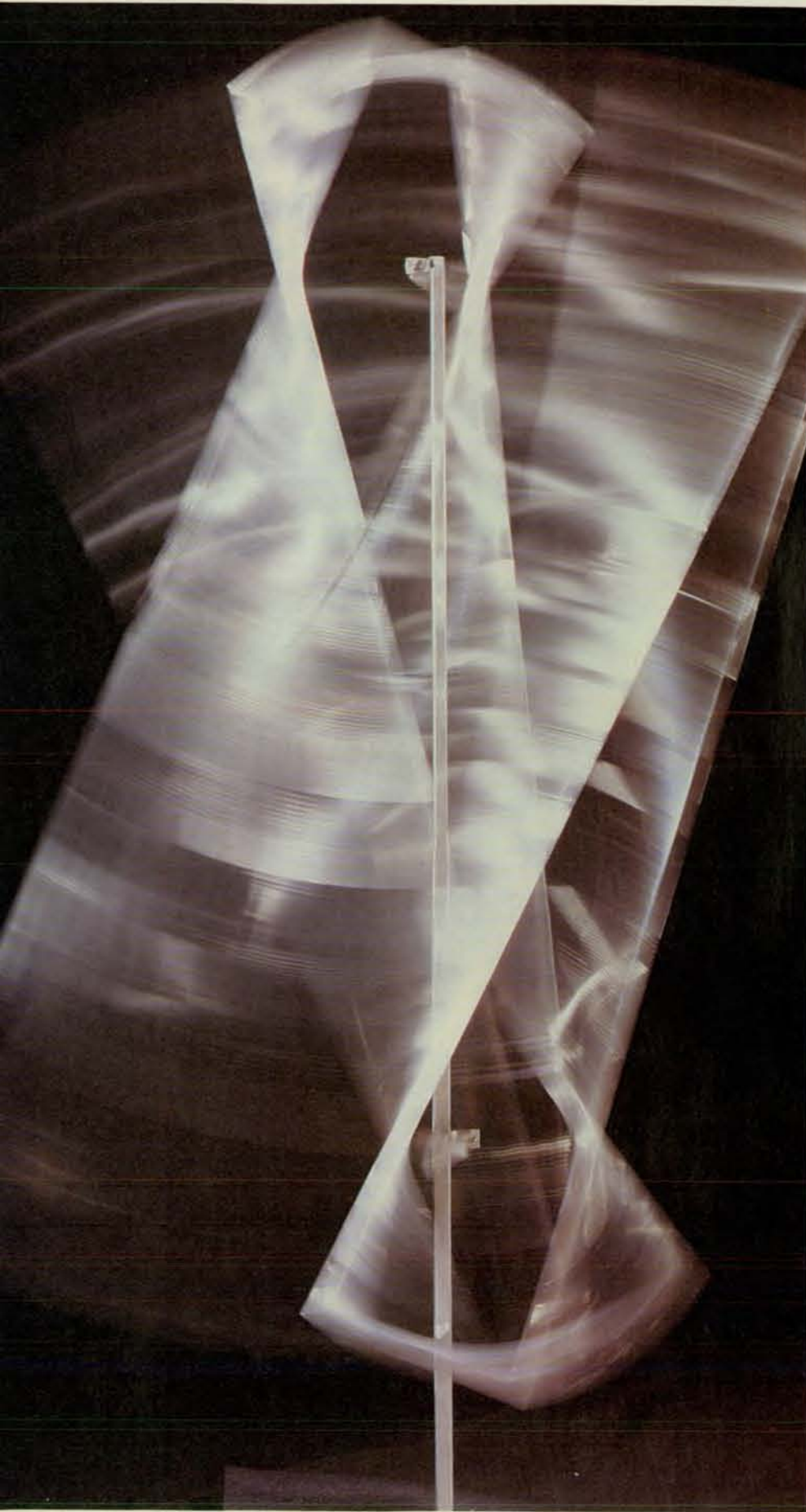
Mosse anche da debolissime correnti d'aria, le sculture dell'autore danzano lievemente grazie a raffinate interazioni di pesi e contrappesi

di George Rickey

L'opera *Breaking Column* ha tre parti controbilanciate che compiono moti conici. I regolari moti periodici di molte sculture cinetiche di Rickey sono dovuti alla precisa calibrazione dei contrappesi di piombo celati in ciascun componente (*a sinistra*). La trave a sezione rettangolare saldata è l'unità strutturale di base di molte opere dell'autore.

Le mie concezioni artistiche si sono sviluppate, come i movimenti delle mie sculture, in modo spontaneo e inatteso. Quando arrivai all'età di 22 anni avevo già lavorato nel campo della moda, avevo insegnato inglese in una scuola di lingue e avevo dato fondo ai miei risparmi di adolescente per frequentare l'Académie Lhote, una scuola di pittura parigina di tendenze cubiste. Ma solo tredici anni dopo, mentre prestavo servizio nello US Army Air Corps durante la seconda guerra mondiale, scoprii nelle mani e nella testa una vocazione alla meccanica.

Mentre lavoravo con attrezzature industriali e militari, cominciai ad accarezzare l'idea di costruire congegni dinamici che non fossero né sculture statiche né composizioni mobili nel senso tradizionale. Al termine della guerra mi accinsi finalmente a trasformare queste idee in una realtà tangibile. Esposi le mie opere per la prima volta nel 1953, all'età di 46 anni.

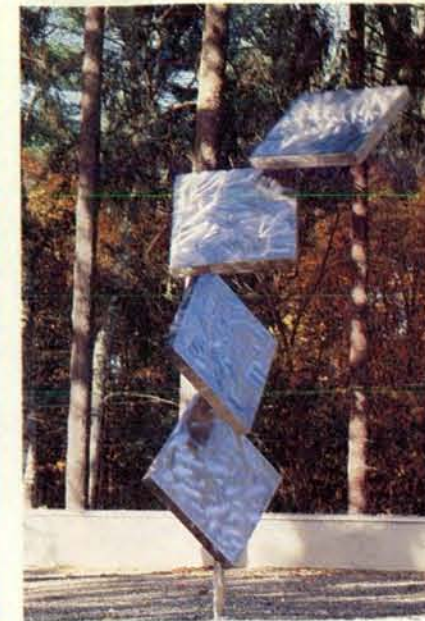
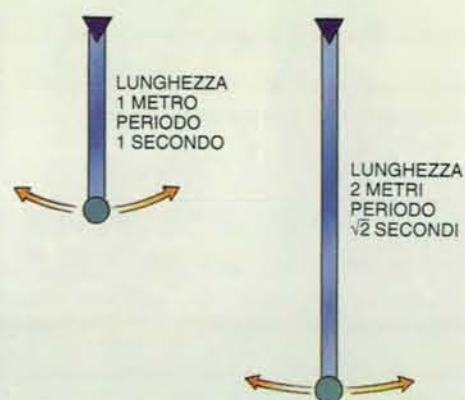
Da allora ho continuato a esplorare le potenzialità estetiche della scultura cinetica. Nelle mie opere più recenti ho incernierato un elemento mobile (spesso una forma geometrica semplice, come un'asta o un cubo) a un altro, ricavandone strutture con due o tre articolazioni. Una distribuzione quasi perfettamente uniforme dei pesi consente loro di rispondere ai minimi refoli di vento tracciando lente evoluzioni, imprevedibili ma espressive.

Al livello più elementare, le mie opere possono essere considerate pendoli modificati. Sono debitore a mio nonno, orologiaio di Athol, nel Massachusetts, per avere spinto la mia fantasia in questa direzione: quando avevo cinque anni, mi fece smontare un orologio svizzero con gli ingranaggi di legno. Lo feci, ma ancora mi dolgo di non essere più riuscito a rimontarlo.

Solo una quarantina d'anni dopo aver fallito la prova cui il nonno mi aveva sottoposto, seppi dell'esistenza dei cosiddetti pendoli composti. A differenza del pendolo semplice di un orologio, un pendolo composto ha un peso sopra e un peso sotto il perno, o fulcro, intorno al quale oscilla l'asta. Nel caso di un pendolo semplice la durata di un'oscillazione, o periodo, dipende solo dalla lunghezza dell'asta, mentre in un pendolo composto il periodo varia al variare della distribuzione dei pesi: quanto maggiore è il peso posto sopra il fulcro, tanto più lenta è l'oscillazione. Quando il prodotto di peso e distanza sopra il fulcro è uguale al prodotto di peso e distanza sotto il fulcro, il periodo tende all'infinito e il pendolo composto diviene instabile. Assumendo una forma differente, un pendolo di questo genere diventa un oggetto familiare: la ruota.

Lavorando con distribuzioni di pesi e distanze che stanno al limite dell'instabilità, riesco a ottenere movimenti lenti, lirici; in una delle mie sculture, mezza oscillazione può durare anche più di venti secondi. Queste costruzioni hanno un'altra caratteristica degna di nota: quando la distribuzione dei pesi si approssima all'instabilità, la forza necessaria a mettere in moto il pendolo diventa quasi nulla. Una brezza appena capace di fare stormire la chioma di un albero può animare una scultura anche di mezza tonnellata. Regolando i pesi e gli equilibri, nonché i tempi, posso avere a disposizione mezzi espressivi paragonabili a quelli che un pittore trae dai colori o un compositore dai suoni.

Ho dedicato gran parte dei miei sforzi al tentativo di uscire dalla strettoia del moto in un solo piano. Una ventina di anni fa, cercando di estendere le potenzialità del movimento espressivo, provai a disporre l'asse di rotazione passante per il fulcro del pendolo non a 90 gradi come al solito, bensì a 45 gradi. La traiettoria che ne risultò non era il normale arco ben noto a chiunque sia stato su un'altalena, bensì un cono. Uscendo dalle due dimensioni, ero entrato in un mondo nuovo.

Spesso questo tipo di rotazione sorprende l'osservatore, poiché fin dall'infanzia siamo abituati soprattutto a movimenti planari. Una tazza che cade dal tavolo, una ruota di bicicletta in rotazione, un pallone in volo verso il portiere, una lenza lanciata verso il fiume e un pianeta orbitante intorno al Sole si muovono tutti più o meno su un piano. I moti conici si presentano nel movimento di una nave, che rulla e beccheggia, si alza e si abbassa e si sposta avanti e indietro. Questi movimenti sono ben noti ai marinai, agli aviatori e a chi soffre il mal di mare. Io ho scoperto che possono essere anche molto belli.

I pendoli composti costituiscono un elemento basilare delle sculture di Rickey. Il periodo di oscillazione di un pendolo semplice dipende esclusivamente dalla sua lunghezza, mentre quello di un pendolo composto è determinato dai prodotti di peso e distanza sopra e sotto il fulcro. Pendoli con pesi opportunamente calibrati possono oscillare lentamente in risposta a forze debolissime. Una fotografia a passo uno dell'opera *Open Triangles One Up One Down (a sinistra)* rivela i suoi lievi movimenti oscillatori.

Questi quadrati collegati sono il risultato di recenti esperimenti di Rickey. Ogni elemento di *Column of Four Squares Excentric Gyratory III* è appesantito e montato su un supporto angolato. L'asta in basso trasmette tutto il peso alla base.

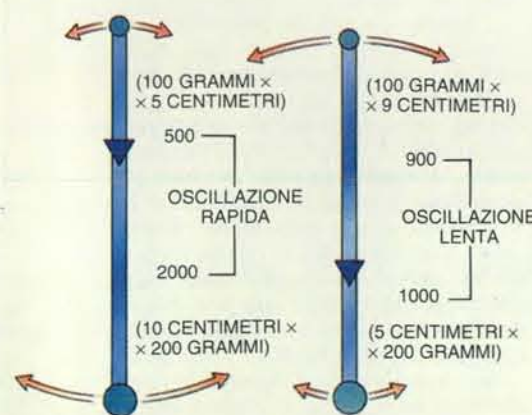
I miei studi sulle costruzioni complesse, dotate di articolazioni, mi hanno portato ad affrontare una moltitudine di problemi tecnici. Ho scoperto che per ottenere risposte sensibili e controllate è necessario che nella scultura la distribuzione delle masse sia accuratamente calibrata. Il peso cumulativo di una successione di parti collegate (le travi rigide, i supporti e i contrappesi) deve equilibrare esattamente il contrappeso dell'unità più interna. Se questa unità più interna non ha dimensioni tali da contenere un contrappeso sufficientemente pesante, la scultura non può essere realizzata secondo il progetto. E, viceversa, lo spazio nel quale inserire i pesi all'interno della massiccia unità interna determina quali possono essere le dimensioni massime dell'unità più esterna.

La rigidità delle travi strutturali è critica quanto la precisione dell'equilibratura: entrambe sono essenziali per costruire opere i cui vaghi movimenti portino lo spettatore in uno stato di sospensione e di attesa. Allo stesso tempo, cerco di fare in modo che le unità più leggere delle mie sculture cinetiche siano indistinguibili da quelle più pesanti. A questo scopo, ho finito con l'affidarmi a una semplice trave a sezione rettangolare, che costruisco saldando a punti due travi d'acciaio a U.

Le mie costruzioni richiedono anche che l'attrito in corrispondenza delle articolazioni sia minimo: una resistenza eccessiva farebbe smorzare prematuramente gli arcani volteggi della scultura. I supporti a coltello, simili a quelli che si impiegano nelle bilance di laboratorio, hanno il vantaggio di presentare un attrito molto ridotto, ma possono deteriorarsi rapidamente o possono venire danneggiati dalla pioggia o da un forte vento. Nelle opere progettate per essere esposte all'aperto mi servo di cuscinetti a sfere, i quali tollerano l'esposizione alle intemperie, ma necessitano di un lubrificante viscoso che può smorzare i movimenti causati dalle brezze più leggere.

Le esigenze dell'equilibratura mi obbligano a costruire una scultura dall'esterno verso l'interno. Quando eseguo un'opera dotata di due parti, comincio con l'equilibrare l'unità esterna, più leggera, poi la unisco alla parte interna, che deve portare un contrappeso sufficiente per equilibrare entrambe le componenti. Infine inserisco l'asse portante, collocato nel baricentro del complesso, in un supporto capace di trasferire al suolo, a una parete o al soffitto il peso di tutta la scultura.

Nel 1985 fui invitato a collaborare, costruendo una scultura monumentale, per le celebrazioni del 750° anniversario della fondazione della città di Berlino, che cadeva nel 1987. La mia scultura doveva essere collocata nella Breitscheidplatz, presso la Gedächtniskirche, nel centro del settore occidentale. Quest'occasione mi spinse a sviluppare ulteriormente le mie tecniche. Avevo già incernierato una componente lineare su un'altra per ricavarne un gomito. Era forse possibile incernierare un cono su un altro cono, come nell'articolazione di un polso o di una spalla? I vertebrati posseggono articolazioni del genere e pensavo che, forse, avrei potuto costruirle anch'io.

Così mi avventurai in questo nuovo territorio: montai una componente angolare su un'altra. Ciascuna parte ruotava intorno a un perno posto a 45 gradi rispetto all'asse maggiore; il complesso era montato su un supporto a zig-zag. Le componenti dondolavano al vento descrivendo coni virtuali ed evitando di pochissimo di urtare il supporto o di collidere tra loro. Calibrai i pesi in modo da ottenere un languido movimento lungo traiettorie che parevano suggerire collisioni inevitabili. Volevo che lo spettatore si chiedesse che cosa sarebbe successo, e quando e dove.

Ulteriori ricerche artistiche sono state applicate alla scultura monumentale. Un'opera eseguita per il 750° anniversario della fondazione di Berlino (*sopra*), che ora si trova presso Stoccarda, ha due parti a zig-zag libere di muoversi su traiettorie coniche. Un'opera successiva, progettata per un teatro di Rotterdam (*sotto*), ha posto problemi tecnici più complessi, perché ciascuno degli elementi ha tre parti mobili.

Disegnare un'opera di tale complessità non era possibile, quindi procedetti per tentativi ed errori, esercitandomi su una versione in miniatura prima di lavorare a grandezza naturale. Ingrandii le unità interne per collocarvi molto piombo. Per ridurre il peso praticai fori nei puntoni di rinforzo. I vincoli fisici della scultura dettarono grandezza e forma complessiva della maggiore delle due unità. Poi, dove le esigenze strutturali lo consentivano, modificai le forme per accrescere l'armonia estetica. La forma definitiva scaturì da un compromesso tra esigenze estetiche e vincoli tecnici.

La scultura ultimata restò a Berlino per due anni, poi venne acquistata dalla Daimler-Benz e installata presso il nuovo edificio dell'amministrazione, a Möhringen, vicino a Stoccarda.

Mentre lavoravo alla scultura di Berlino, un architetto, incaricato di progettare un nuovo teatro a Rotterdam, mi propose di pensare a una scultura per la facciata. Mi mandò i progetti e alcune informazioni sul sito. Il teatro era circondato da edifici a sei piani per uso abitativo. La facciata era dominata da tre quadrati di 10 per 10 metri formati da travi con sezione ad H, che sporgevano di cinque metri dalla parete. Servendomi di stecche di legno, costruii il modello di un gruppo di unità lunghe e sottili, collegate da perni posti a 45 gradi l'uno rispetto all'altro, in modo da movimentare la facciata del teatro, alta quindici metri. Inoltre queste unità dovevano allontanarsi dalla parete per riempire lo spazio racchiuso dalle travi ad H. Ciascun perno di collegamento era incernierato sull'elemento vicino in modo da creare movimenti conici su tre livelli.

Nella versione definitiva la scultura consiste in tre colonne, ciascuna delle quali comprende quattro segmenti collegati. Il primo è un corto e robusto plinto fisso, imbullonato a una trave ad H di ciascuno dei quadrati della facciata. Ciascun supporto sostiene una colonna alta sei metri, con una sezione di 35 per 18 centimetri. La colonna centrale s'innalza dal supporto, mentre quelle laterali pendono dai loro supporti, che sono collocati in alto, evocando in un certo modo la posizione tradizionale delle tre Grazie, due delle quali guardano in avanti e l'altra all'indietro.

Ultimata l'opera di Rotterdam, ho continuato a esplorare le potenzialità delle strutture a più componenti collegate da perni a 90 e a 45 gradi. Inoltre ho ripreso alcune sperimentazioni cinetiche più semplici. Per esempio ho montato un gruppo di quattro quadrati in modo che descrivessero un movimento conico verso l'esterno (e non verso l'alto), come i petali di un fiore. Questo motivo floreale mi ha condotto in due direzioni: verso colonne comprendenti una pila di quadrati rotanti e verso una composizione a grappolo di cubi rotanti.

Avevo già tentato di montare un quadrato sull'altro, facendoli ruotare intorno a un perno a 45 gradi; ora invece collocai questo complesso su un altro quadrato, anch'esso a 45 gradi, e poi i tre quadrati su un quarto che si muoveva intorno a un asse verticale. Ciascun quadrato si muove lungo una traiettoria conica diretta verso l'esterno rispetto a quello sottostante. Il risultato è una forma colonnare che sembra spezzarsi e poi ricomporsi, disintegrandosi e reintegrandosi al ruotare dei quadrati. Un'opera di questo tipo fu presentata all'esposizione di Katonah; una versione più grande si trova presso il nuovo municipio di Tokyo.

La natura dei supporti determina la durata e la libertà di movimento delle sculture cinetiche. Nelle opere progettate per interni si possono usare sottili supporti a coltello (*nelle tre foto in basso*), mentre le opere da esporre all'esterno richiedono sensibili cuscinetti a sfera ricoperti (*qui sotto*). L'autore studia i movimenti delle proprie sculture ricorrendo a modellini in scala (*a destra*).

Dopo gli esperimenti con questi quadrati rotanti, cominciai a domandarmi che cosa sarebbe accaduto se avessi trasformato i quadrati in cubi. Mi resi conto che, su un asse diagonale da vertice a vertice, essi avrebbero avuto abbastanza spazio per ruotare senza urtarsi. Ma un cubo era tanto aerodinamico da essere messo in movimento da una lieve brezza? Costruii un modello di prova, il quale mi rispose lentamente di sì. Perciò a Katonah fui in grado di esporre quattro cubi a grappolo rotanti. Una versione più grande e perfezionata di questo grappolo di cubi fu installata nell'autunno 1992 presso l'ala orientale della National Gallery di Washington.

Ora sono impegnato in un altro progetto di grandi dimensioni, una scultura sospesa per un atrio di otto piani a Berlino, che installerò l'estate prossima. Tuttavia so che le mie esplorazioni artistiche si avviano alla fine. L'età rallenta la mano, la memoria e lo slancio, e comincio a sentire il desiderio di ritirarmi e di ricordare, ma anche di sbirciare dietro l'angolo per vedere che cosa mi resta da fare. Mi vengono in mente le parole che Shakespeare fa pronunciare a Riccardo II in prigione: «E nondimeno io voglio riuscirvi, in un modo o nell'altro. Cercherò di far del mio cervello la femmina e del mio spirito il maschio, e tutt'e due metteranno al mondo una generazione di pensieri che, a loro volta, ne figlieranno altri, e tali pensieri popoleranno questo piccolo mondo...». Se la Musa mi soccorrerà un po'.

BIBLIOGRAFIA

RICKEY GEORGE, *Constructivism: Origins and Evolution*, George Braziller, 1967.

ROSENTHAL NAN, *George Rickey*, Harry N. Abrams, 1977.

RICKEY G., ROSENTHAL N. e SNYDER J., *George Rickey: Art of Movement*, Katonah Museum of Art, 1991.

Infezioni batteriche e barriera ematoencefalica

Studiando i meccanismi che consentono ai batteri di invadere l'encefalo, si è messa a punto una terapia combinata per la meningite e si sono proposte nuove strategie contro altre malattie che colpiscono il cervello

di Elaine Tuomanen

Poche volte, in medicina, si può provare una frustrazione paragonabile a quella del neurologo che non riesce a far giungere un farmaco efficace fino al sito di un'infezione o di un tumore. Il cervello è il solo organo che rimanga inaccessibile alla chemioterapia. Mentre le pareti dei vasi sanguigni che irrorano l'organismo presentano minuscole aperture, come un muro pieno di feritoie, quelle dei vasi cerebrali assomigliano piuttosto a un muro perfettamente liscio e compatto. La cosiddetta barriera ematoencefalica permette il passaggio del glucosio e di poche altre sostanze specifiche, ma esclude tutti gli intrusi, che siano ostili o meno.

Come in battaglia, però, una spia può infiltrarsi al di là di una linea che nessun esercito può oltrepassare. Certi batteri riescono in effetti a superare furtivamente la barriera ematoencefalica e a infettare il liquido cerebrospinale. Si tratta degli agenti della meningite batterica, che negli Stati Uniti è l'infezione infantile più temuta e nel mondo è una delle principali cause di morte per adulti e bambini. Negli Stati Uniti essa colpisce circa 50 000 persone all'anno; fino a tempi recenti la mortalità dovuta a questa malattia arrivava al 30 per cento e più di metà dei sopravvissuti presentava complicanze infiammatorie del sistema nervoso centrale. La situazione è ancora più sconfortante per i pazienti il cui sistema immunitario è già stato compromesso dall'AIDS. I miei colleghi e io abbiamo fatto numerose scoperte sul meccanismo con cui i batteri causano la meningite, e il nostro lavoro ha portato a nuovi trattamenti terapeutici che migliorano notevolmente le possibilità di sopravvivenza. Potremmo addirittura essere sul punto di comprendere come le «spie» batteriche superino la barriera ematoencefalica stessa. Se così fosse, avremmo a portata di mano l'allettante possibilità di poter sfruttare il loro segreto per infiltrare farmaci attraverso la barriera ematoencefalica, esponendo per la prima volta i tumori cerebrali e la malattia di Alzheimer a tutto l'arsenale della moderna medicina.

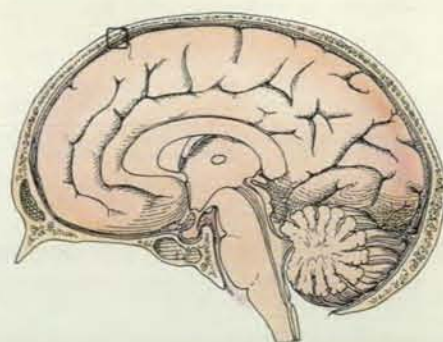
Tre sono i microrganismi che possono causare meningite acuta. *Haemophilus influenzae* (che non ha nulla a che vedere con il virus dell'influenza vera e propria) è stato il più comune fino al 1990, quando la somministrazione diffusa di un nuovo vaccino ha iniziato a debellarlo. *Neisseria meningitidis*, o meningococco, continua a causare epidemie in Africa, dove la distribuzione del vaccino è insufficiente. Il più insidioso è *Streptococcus pneumoniae*, o pneumococco, che provoca anche polmonite e infezioni dell'orecchio medio. La meningite pneumococcica uccideva circa il 30 per cento dei bambini infettati, per quanto intensive fossero le cure, fino a un paio d'anni fa, quando i miei colleghi e io trovammo un metodo per ridurre la mortalità ben al di sotto del 5 per cento. Le tre forme di meningite batterica seguono tutte lo stesso decorso. Il paziente manifesta febbre, irritabilità e sonnolenza, che sono sintomi caratteristici dell'influenza virale; compare poi l'infiammazione cerebrale che può causare convulsioni, coma e, in molti casi, morte. Il medico deve perciò essere attento ai sintomi più sfuggenti della meningite, in particolare la rigidità nucale. Il semplice sospetto della malattia deve spingere a una terapia di emergenza che preveda la somministrazione endovenosa di antibiotici. Quando si iniettano antibiotici in dosi molto elevate, una piccola quantità riesce di solito a superare la barriera ematoencefalica e a uccidere la maggior parte dei batteri. Ma il medico deve ricorrere a un vero e proprio cocktail di farmaci: non può infatti permettersi il lusso di attendere un giorno intero per identificare il microrganismo invasore e scegliere l'antibiotico specifico.

Nelle successive 24 ore, due domande si impongono: l'antibiotico sta effettivamente uccidendo i batteri? E il paziente riuscirà a sopravvivere? Abbiamo scoperto che queste domande sono tragicamente correlate. Antibiotici come la penicillina debellano totalmente i tre agenti della meningite, eppure un bambino su tre muore. Questo problema assillava tutti i membri del Cooperative Meningitis Study Group, un'associazione informale che organizza incontri di studio ad anni alterni. Se l'infezione è curabile, perché molti pazienti non sopravvivono?

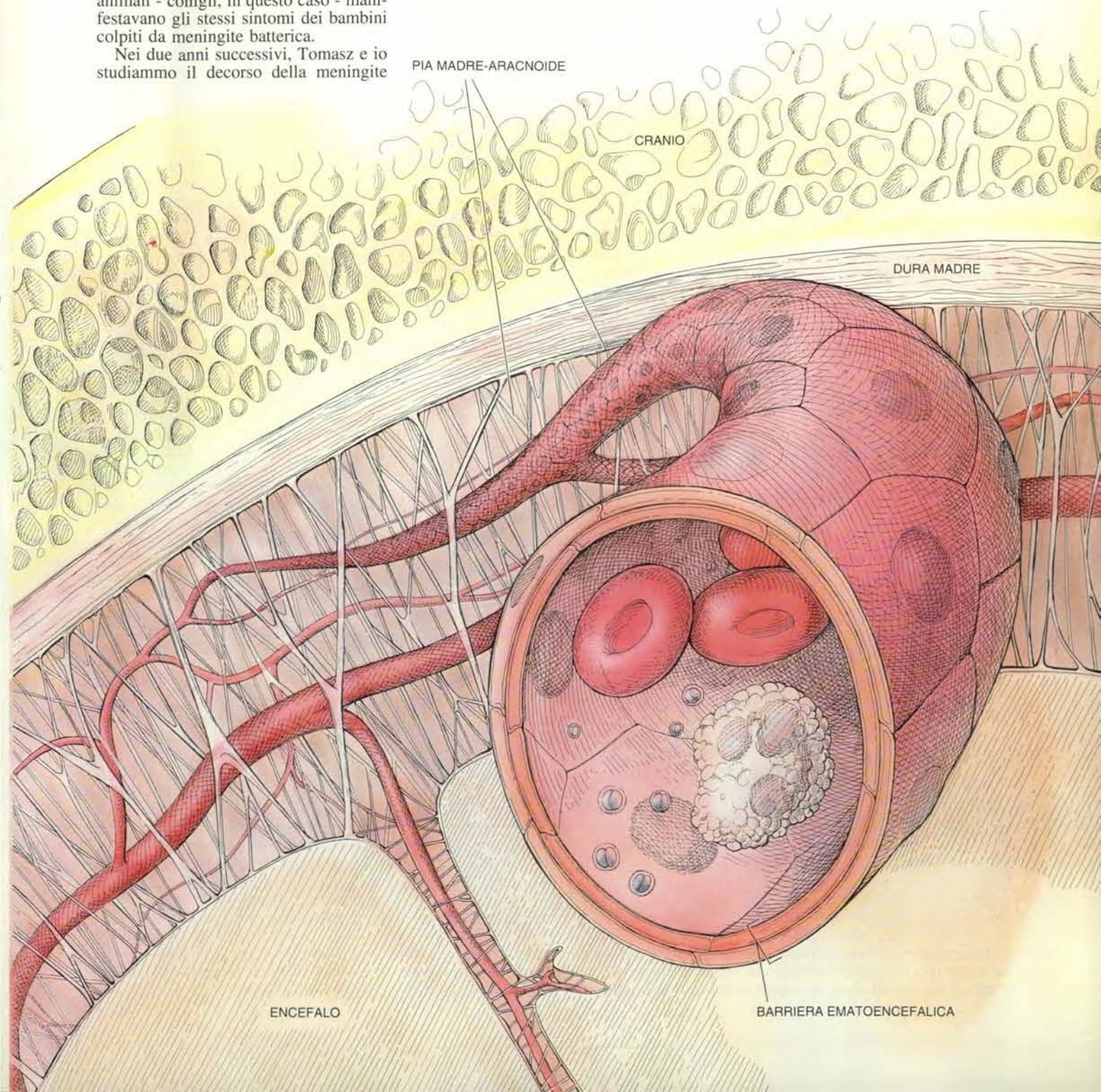
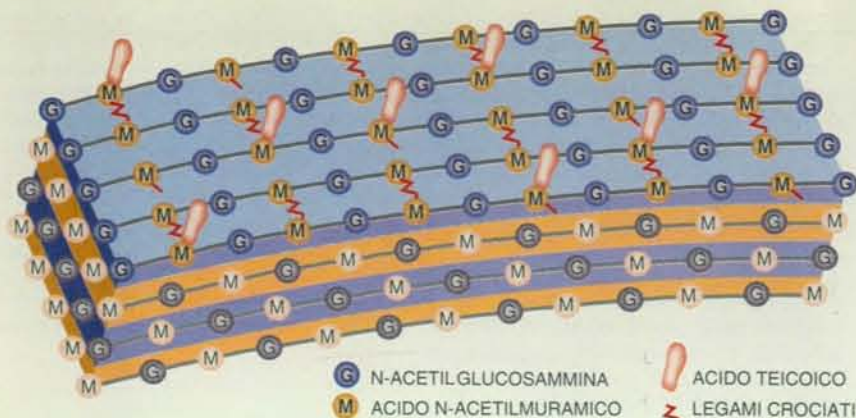
Ricordo il giorno dell'estate 1984 in cui cominciai a comprendere come si potesse giungere alla risposta. Il mio collega Alexander Tomasz mi chiamò dalla Svizzera, dove stava trascorrendo un periodo sabbatico presso i Laboratori di ricerca della Ciba-Geigy a Basilea. Il suo studio consisteva nel determinare, in animali di laboratorio, il meccanismo di uccisione dei batteri da parte degli antibiotici. Egli aveva accidentalmente esposto animali da esperimento a pneumococchi uccisi ed essi si erano ugualmente ammalati. Sembrava dunque che i batteri uccisi fossero dannosi quanto quelli vivi. Come si spiegava questo fatto?

Quella settimana volai a Basilea per aiutare Tomasz a ripetere l'esperimento; i risultati furono gli stessi. In quanto pediatra, notai immediatamente che gli animali - conigli, in questo caso - manifestavano gli stessi sintomi dei bambini colpiti da meningite batterica.

Nei due anni successivi, Tomasz e io studiammo il decorso della meningite servendoci di conigli per definire un modello della malattia, secondo una procedura messa a punto all'inizio degli anni ottanta da Ralph G. Dacey, Jr., e Merle A. Sande dell'Università della Virginia. Le nostre ricerche dimostrarono che i batteri uccisi non provocano danni diretti, ma sollecitano l'organismo a un parossismo autodistruttivo.

I frammenti dei batteri uccisi che stimolano il sistema immunitario provengono dalla parete cellulare, un esoscheletro rigido costituito da catene di polisaccaridi unite da legami crociati; senza la parete il batterio esplode. Tomasz aveva dimostrato che la maggior parte dei batteri, incluso lo pneumococco, morendo si disintegra in un processo che comporta la distruzione della parete cellulare. La penicillina sfrutta questa tendenza suicida inducendo la lisi, un processo che coinvolge gli stessi enzimi batterici, i quali di norma mantengono la parete cellulare in perfette condizioni. La penicillina distrugge i controlli che limitano l'azione degli enzimi, cosicché questi finiscono per dissolvere la parete cellulare. A quasi 50 anni dalla scoperta della penicillina, non si è ancora risolto

La barriera ematoencefalica, un rivestimento compatto dei vasi sanguigni cerebrali, impedisce il passaggio alla maggior parte delle sostanze. La meningite si manifesta quando i batteri superano la barriera, di solito impenetrabile, e si moltiplicano nel liquido cerebrospinale.

PIA MADRE-ARACNOIDE
CRANIO
DURA MADRE
ENCEFALO
BARRIERA EMATOENCEFALICA

N-ACETILGLUCOSAMMINA
ACIDO N-ACETILMURAMICO
ACIDO TEICOICO
LEGAMI CROCIATI

La parete di un batterio consiste di una singola molecola con legami crociati. La penicillina frantuma la parete in milioni di frammenti, formati da uno dei circa 20 elementi che la costituiscono e che provocano un diverso sintomo dell'infezione.

il mistero della sua azione. Avevamo però il sospetto che questa morte «esplosiva» potesse danneggiare l'organismo pur estinguendo l'infezione.

Il nostro modello sperimentale indicava che all'inizio, quando i batteri si moltiplicano nel liquido cerebrospinale, l'animale non presenta sintomi. Tuttavia, quando la concentrazione di batteri sale a 100 000 per millilitro, è come se un incantesimo si spezzasse. Si manifestano febbre e infiammazione, mentre le difese dell'organismo chiamano improvvisamente in causa sostanze simili a ormoni, le citochine (interleuchina 1 e fattore di necrosi tumorale). Le citochine provengono dal lato cerebrale della barriera ematoencefalica e aprono quest'ultima per permettere l'ingresso dei globuli bianchi. I batteri uccisi inducono la stessa successione di eventi alla stessa concentrazione, vale a dire 100 000 per millilitro. In un paziente non curato, dunque, sia i resti batterici sia i batteri vivi favoriscono il ciclo dell'infiammazione e le sue conseguenze.

In che modo i batteri uccisi possono far sì che l'organismo si autoinfligga danni di questa portata? È possibile che il sistema immunitario confonda invasori vivi e morti? Per scoprirlo, abbiamo inoculato in animali da esperimento appena un migliaio di pneumococchi vivi, somministrando subito dopo penicillina. Poi abbiamo atteso per vedere se i batteri esplosi avrebbero indotto i globuli bianchi a invadere il cervello. Trenta minuti dopo, in effetti, il liquido cerebrospinale pullulava di globuli bianchi.

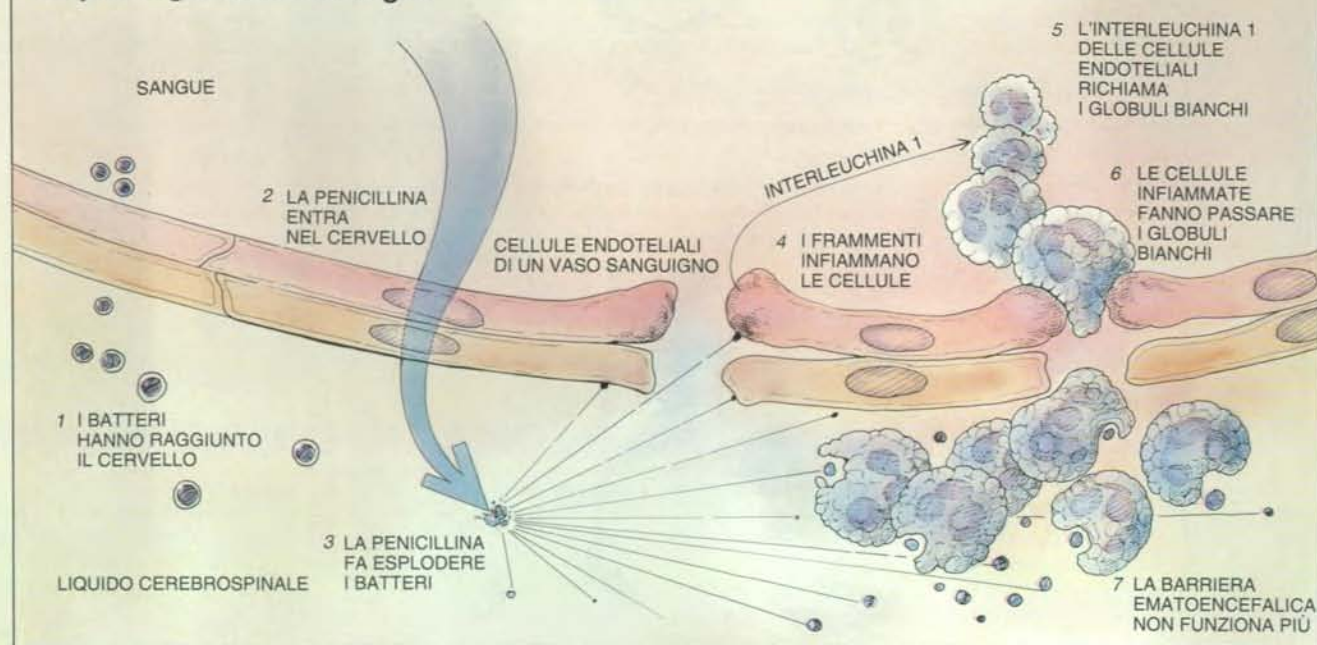
Poco dopo questi esperimenti, Martin G. Täuber, Jay H. Tureen e Sande eseguirono studi analoghi al San Francisco General Hospital. Le loro ricerche dimostrarono che nelle prime ore di terapia antibiotica non solo aumentava la proliferazione infiammatoria, ma l'accumulo di essudato e la pressione del liquido cerebrospinale raggiungevano livelli pericolosi. Un'infiammazione così violenta può danneggiare irreparabilmente la struttura cellulare del cervello e dei tessuti di supporto.

Il nostro sospetto era stato confermato. Quando un batterio esplode, le difese dell'organismo scambiano la pioggia di frammenti per un'improvvisa proliferazione batterica e reagiscono di conseguenza. I resti della parete cellulare stimolano le difese, attivando le citochine e i fattori iniziatori della coagulazione. Questi processi preparano sulla parete del vaso una «piattaforma» a cui i globuli bianchi aderiscono in attesa di riuscire a infiltrarsi al di là della barriera ematoencefalica. Una volta attestatisi nel cervello, i globuli bianchi stimolano ulteriormente la produzione di citochine e accelerano così la distruzione della barriera. Altri globuli bianchi migrano nel cervello, aggravando l'infiammazione, l'edema e la liberazione di sottoprodotti tossici della reazione immunitaria.

In poche parole, il modello animale ci aveva fornito una lezione inattesa: la terapia antibiotica aggrava la meningite prima di alleviarla. Grazie a questa osservazione, il gruppo diretto da George H. McCracken, Jr., del Southwestern Medical Center della Università del Texas a Dallas riuscì a dimostrare che anche nei bambini sottoposti a terapia per la meningite si hanno analoghi effetti dannosi. Tutti i pazienti morirebbero senza una terapia antibiotica, ma alcuni non sopportano la terapia; quelli che sopravvivono perdono numerosi neuroni e, in molti di essi, permangono deficit di udito, movimento o apprendimento.

Per salvare il paziente dagli effetti negativi della terapia antibiotica, occorre trovare un modo per attenuare le difese immunitarie mentre gli antibiotici svolgono la loro azione. Per determinare quali sistemi di difesa debbano essere soppressi e quali no, occorrerebbe una conoscenza dettagliata di tutti gli eventi che concorrono alla genesi della malattia. Queste ricerche sono ora condotte dal Cooperative Meningitis Study Group, ma per avere i risultati definitivi occorreranno anni. Nel frattempo, stiamo cercando una soluzione provvisoria.

Il desametasone, il più potente antinfiammatorio steroideo, annulla quasi del tutto il processo infiammatorio. Eravamo quindi convinti che avrebbe quasi certamente impedito alle difese dell'ospite di reagire in maniera incontrollata durante le prime, critiche ore di terapia antibiotica. Tuttavia, benché anni di esperienza clinica avessero dimostrato che il farmaco è efficace contro le in-

Gli pneumococchi intatti hanno uno strato capsulare filamentoso sulla parete cellulare (*a sinistra*). La penicillina distrugge la parete dei batteri facendoli esplodere e disseminando di resti di parete cellulare il campo del microscopio (*a destra*).

fiammazioni (per esempio contro l'artrite), la sua efficacia nelle infezioni non era certa. Inoltre alcuni medici continuavano a opporsi all'idea di reprimere le difese dell'organismo proprio quando l'infezione era più intensa. Per questi motivi fummo costretti a sperimentare il farmaco in animali di laboratorio prima di passare ai bambini.

Decidemmo di mettere alla prova sia farmaci steroidei sia composti a essi correlati, egualmente potenti ma più selettivi, gli antinfiammatori non steroidei. Per fortuna i risultati furono rapidi e inequivocabili. Lavorando con Oto Zak dei Laboratori di ricerca della Ciba-Geigy, dimostrammo che la somministrazione di entrambi i tipi di farmaci non solo bloccava la risposta infiammatoria all'uccisione dei batteri, ma favoriva anche la sopravvivenza. In alcuni trattamenti con composti non steroidei, la mortalità scendeva quasi a zero.

Purtroppo la Food and Drug Administration non ha approvato l'uso di antinfiammatori non steroidei per la cura di malattie infettive nei bambini. Tuttavia si potevano usare gli steroidi, e nel 1987 il gruppo di McCracken a Dallas iniziò una sperimentazione clinica che combinava steroidi e antibiotici. Il trattamento diminuì l'incidenza dei casi di sordità dal 15 al 3 per cento e abbreviò da 5 a 1,6 giorni la durata della febbre. Questi risultati furono dapprima accolti con scetticismo, ma da allora sono stati confermati da molti gruppi di ricerca in tutto il mondo e ora si riconosce la validità del nostro modello.

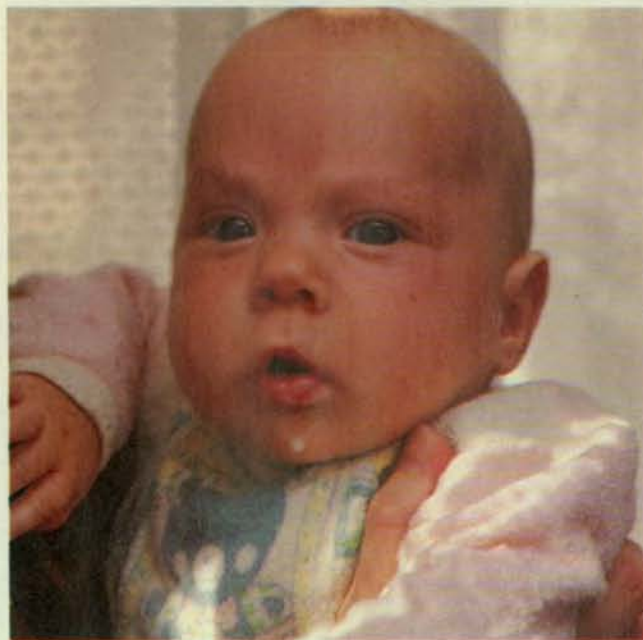
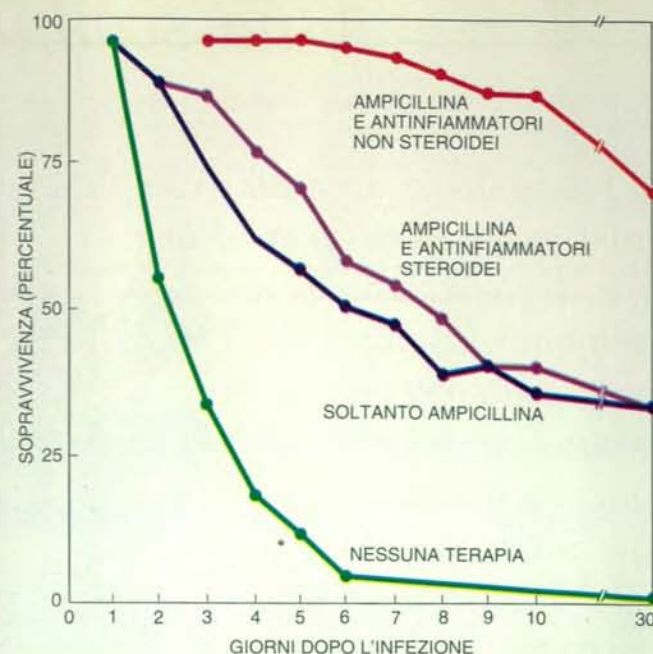
Nel 1990 la American Academy of Pediatrics accolse l'uso di farmaci steroidei combinati con antibiotici come terapia d'elezione per la meningite infantile. W. Michael Scheld e colleghi dell'Università della Virginia stanno ora cercando di dimostrare i vantaggi dell'impiego di steroidi nella meningite in sperimentazioni cliniche condotte sugli adulti. Ma gli steroidi hanno effetti collaterali significativi, come emorragie gastrointestinali, che possono rivelarsi gravi in un bambino già provato dall'infezione. Ritengo che si possano trovare terapie molto più efficaci.

Il problema più grave è il modo brutale con cui la maggior parte degli antibiotici uccide i batteri. Si vorrebbe quindi disporre di un «killer» più raffinato. Una nuova famiglia di antibiotici beta-lattamici, come l'imipenem, si avvicina a questo ideale. Sebbene anche queste sostanze spezzino la parete cellulare, i frammenti sono meno nocivi di quelli prodotti dall'azione della penicillina. Un'altra strategia è quella di neutralizzare i frammenti della parete cellulare, e il modo più sicuro e specifico per farlo è mascherare i frammenti con anticorpi su misura. Il mascheramento dovrebbe impedire alle difese immunitarie di riconoscere i frammenti come corpi estranei. Ma si presenta di nuovo il problema della barriera ematoencefalica, che impedisce il passaggio degli anticorpi.

Se vogliamo far arrivare all'encefalo gli anticorpi, dobbiamo trovare un modo per far breccia nella barriera, pur senza danneggiare le cellule nervose, il cui ambiente normale è completamente diverso da quello del siero ematico. Non sarebbe addirittura meglio trovare un anticorpo che possa impedire ai globuli bianchi di accedere al cervello?

Per conseguire questo risultato ho tentato di utilizzare l'anticorpo anti-CD18, che impedisce ai globuli bianchi di aderire alle pareti dei vasi. Con questo trattamento, le cellule non dovrebbero trovare il modo di migrare nel cervello per alcune ore, quanto basta perché la fase più violenta dell'attività antibiotica abbia termine. Per verificare la mia ipotesi, ho inoculato in animali di laboratorio prima anti-CD18 e poi penicillina.

Gli eventi delle ore successive sono tra i più vivi nella mia memoria. Gli animali infettati cui era stato somministrato l'anticorpo insieme con gli antibiotici tornarono rapidamente vigili e ripresero la normale attività come se non fossero stati ammalati; la percentuale di sopravvivenza fu del 100 per cento. I benefici causati dall'aver impedito la distruzione della barriera ematoencefalica da parte dei globuli bianchi furono evidentissimi. Si prevede che questo anticorpo sarà disponibile fra alcuni mesi per sperimentazioni cliniche sull'uomo.

La meningite è stata piegata, ma non sconfitta una volta per tutte. A questo scopo, ci proponiamo di identificare gli aspetti più nocivi del decorso della malattia per poi controbattere specificamente ciascuno di essi. Sapendo che i frammenti di batteri esplosi, soprattutto quelli della parete cellulare, contengono le informazioni necessarie per provocare tutti i segni e i sintomi della malattia, riteniamo che ciascun frammento contribuisca a causare uno o più sintomi. Dato che ogni specie batterica ha una propria struttura della parete cellulare e un proprio insieme di enzimi che possono distruggerla, le varie infezioni batteriche

La terapia proposta dall'autrice prima inibisce le difese immunitarie, poi uccide i batteri. Questo bambino è perfettamente guarito dalla meningite. Pare che in tal modo la mortalità sia scesa dal 30 a meno del 5 per cento. L'indice si abbasserà ancora se gli antinfiammatori non steroidei, sinora sperimentati su animali, saranno approvati per uso pediatrico.

differiscono in maniera prevedibile. È necessario quindi conoscere a fondo la struttura della parete in ogni specie.

Per classificare tutte le subcomponenti delle pareti che avvolgono i principali, agenti patogeni dell'uomo ci serviamo di una tecnica messa a punto in Germania. La parete di ciascun microrganismo invasore contiene una ventina di questi mattoni, che il nostro e altri gruppi di ricerca stanno tentando di caratterizzare. Lavori recenti hanno dimostrato che, come previsto, i diversi frammenti hanno attività biologica differente.

Sembra che questa regola valga anche per infezioni che colpiscono altre regioni dell'organismo. Abbiamo dimostrato che i frammenti di parete cellulare causano infiammazioni nei polmoni e nell'orecchio. Raoul S. Rosenthal della Indiana University School of Medicine e William E. Goldman della Washington University hanno dimostrato che un frammento, comune a molte cellule batteriche, può provocare artrite e uccidere le cellule epiteliali delle vie respiratorie. James M. Krueger del Center for Health Sciences dell'Università del Tennessee ha stabilito che un altro frammento induce il sopore, e probabilmente anche il coma, che si accompagna alle infezioni più gravi. Forse ulteriori studi chiariranno se piccole dosi di questa sostanza possano fungere da sonnifero naturale.

Le nostre scoperte sugli effetti che i frammenti di parete cellulare hanno su molte funzioni dell'organismo ci hanno spinti a collaborare con la Alkermes, Inc., di Cambridge, nel Massachusetts, per rovesciare le parti in questa storia di spie. Il nostro piano consiste nel volgere a nostro vantaggio le capacità dell'invasore. Se dagli pneumococchi uccisi si pensa di ottenere un farmaco, si può sperare di ricavare qualche vantaggio anche dalle altre specie batteriche?

Ho scoperto una possibilità particolarmente interessante in *Bordetella pertussis*, l'agente eziologico della pertosse. Esso inganna l'organismo producendo segnali chimici che interferiscono con la capacità dei globuli bianchi di riconoscere una chiamata d'emergenza. La parte bioattiva di questo batterio potrebbe essere sfruttata per un agente antinfiammatorio altamente selettivo in grado di curare infiammazioni del cervello, delle articolazioni, dell'occhio e di altri siti dove il tessuto circostante deve essere protetto sia dall'attività del sistema immunitario sia da quella degli invasori.

Queste ipotesi ci conducono al nostro obiettivo più interessante: il segnale batterico che permette ai microrganismi di attraversare la barriera ematoencefalica e provocare la malattia. Tomasz e io stiamo cercando la chiave per aprire la barriera ematoencefalica spezzando la parete cellulare dello pneumococco con gli stessi enzimi che il batterio impiega per la lisi. I frammenti vengono poi introdotti in un dispositivo per cromatografia liquida ad alta pressione, che li separa in base alla composizione chimica. Inoculiamo poi ciascun tipo di frammento in un coniglio e controlliamo il risultato iniettando nell'animale molecole di un tracciante che normalmente non può entrare nel cervello. Se il tracciante trova una strada per accedervi, è chiaro che la barriera è stata aperta.

Al momento di andare in stampa, riteniamo di aver trovato una possibile molecola: un glicopeptide, ossia una proteina legata a uno zucchero. Potrebbe però essere solo una delle chiavi che aprono la barriera ematoencefalica, e ne stiamo cercando altre.

La selezione naturale ha fornito alle cellule batteriche molte parole d'ordine, ognuna delle quali induce un processo normale nell'ospite umano e può quindi essere sfruttata a scopo terapeutico. Del resto l'umanità ha già imparato a servirsi di molte tossine naturali, come il curaro. Siamo riusciti a controllare la risposta violenta del sistema immunitario alla morte dei batteri; ora stiamo interrogando i sospetti agitatori. La spia silenziosa che si infiltra nelle difese del cervello non potrà sfuggirci a lungo.

BIBLIOGRAFIA

TUOMANEN E., *Partner Drugs: A New Outlook for Bacterial Meningitis* in «Annals of Internal Medicine», 109, n. 9, 1 novembre 1988.

SAEZ-LLORENS X., RAMILO O., MUSTAFA M. M., MERTSOLA J. e McCRACKEN G. H., Jr., *Molecular Pathophysiology of Bacterial Meningitis: Current Concepts and Therapeutic Implications* in «Journal of Pediatrics», 116, n. 5, maggio 1990.

QUAGLIARELLO V. e SCHELD W. M., *Bacterial Meningitis: Pathogenesis, Pathophysiology and Progress* in «New England Journal of Medicine», 327, 17 settembre 1992.

Come dovrebbero ragionare i chimici?

Uno scienziato si chiede quale debba essere l'obiettivo principale della ricerca chimica: se ottenere artificialmente molecole che esistono in natura oppure creare nuove strutture di valore prevalentemente estetico

di Roald Hoffmann

In Vaticano è conservato l'affresco di Raffaello intitolato *La scuola di Atene*. In esso sono raffigurati Platone e Aristotele che incedono verso chi guarda. Platone accenna al cielo con un dito, Aristotele invece ha la mano aperta rivolta in basso, verso la terra. Il messaggio sintetizza le concezioni dei due filosofi: mentre Platone aveva formulato una prototeoria geometrica a proposito della chimica della materia, Aristotele descriveva in dettaglio come estrarre la porpora di Tiro dal murice. Mentre Platone si dedicava alla ricerca dell'idea, Aristotele guardava alla natura.

È curioso il fatto che la chimica moderna debba risolvere lo stesso dilemma rappresentato nell'affresco di Raffaello. Ci si domanda se sia meglio seguire il gesto di Aristotele o quello di Platone. La natura è realmente una fonte prodiga di idee per realizzare nuovi materiali, come sostiene qualcuno? Possiamo sperare, per esempio, di ottenere materiali compositi migliori imitando la microstruttura di una piuma o della tela di un ragno? O i chimici farebbero meglio a ispirarsi a forme matematicamente perfette, come icosaedri e palloni da calcio? Oppure dovrebbero cercare a caso, sperando nella buona sorte?

Per alcuni, la distinzione tra ciò che è naturale e ciò che è artificiale è arbitraria; costoro potrebbero sostenere che, dato che uomini e donne sono evidentemente esseri naturali, naturali devono essere anche tutte le trasformazioni da loro realizzate. Questa visione è comprensibile e ha una rispettabilissima tradizione, ma sottovaluta una distinzione che appare chiara al buon senso. Per questo motivo ho deciso di non adottarla e ho preferito fare una distinzione tra le azioni, generalmente premeditate, degli esseri umani e quelle degli animali, delle piante e del mondo inanimato che ci circonda. Un tramonto è un evento naturale; una fabbrica di acido solforico non lo è. Ma come definire il miliardo e trecento milioni di capi di allevamento sparsi per il mondo? Questi animali possono essere considerati per la maggior parte sia naturali sia artificiali: sono infatti il frutto di incroci controllati dall'uomo.

Le molecole presenti in natura sulla Terra sono il risultato di miliardi di anni di trasformazioni in cui si sono raffreddate le rocce, si sono formati gli oceani e l'involucro gassoso del pianeta e si è evoluta la vita. Il numero di molecole naturali è immenso: probabilmente le centinaia di migliaia isolate, purificate e identificate finora non sono che una piccola parte. La maggior parte dei composti artificiali è stata creata negli ultimi tre secoli. I chimici hanno aggiunto circa 15 milioni di molecole ben caratterizzate al munifico repertorio naturale.

A ogni oggetto di questo mondo, animato o inanimato, corrisponde una struttura. Al primo livello ci sono le molecole, che sono raggruppamenti stabili di atomi associati fra loro. L'acqua può essere pura come la si distilla in laboratorio, o un po' sporca e acida come nella neve, o associata alle molecole che costituiscono le proteine del nostro organismo, ma si tratta sempre di H_2O. Quando la chimica era ancora alla ricerca di un consenso, l'opinione generale era che il mondo animato fosse ben distinto da quello inanimato. Friedrich Wöhler dimostrò che questi due mondi non erano separati quando, nel 1828, sintetizzò l'urea organica a partire da cianato d'argento e cloruro di ammonio.

Come vengono prodotte le molecole in natura, si tratti della penicillina di una muffa o della porpora di un murice? E come si fa, nei laboratori scintillanti di provette, a ottenere chilometri quadrati di involucri trasparenti per alimenti e miliardi di compresse di aspirina? Attraverso un medesimo processo: la sintesi.

La chimica è la scienza delle molecole e delle loro trasformazioni. Sia che si tratti di una trasformazione spontanea o di una guidata dall'uomo, il risultato della trasformazione $A \rightarrow B$ è una nuova sostanza. La sintesi chimica, che realizza il nuovo, è evidentemente un atto creativo. È un'affermazione di umanità, esattamente quanto lo sono la composizione di una poesia o la realizzazione di una nuova democrazia. Ma come ogni atto creativo, comporta dei rischi. Un nuovo sedativo può essere efficace, ma può anche indurre malformazioni fetali. Una poesia di Heberto Padilla può essere giudicata controrivoluzionaria dall'«apparato» di Cuba. In Russia c'è chi diffida ancora della democrazia.

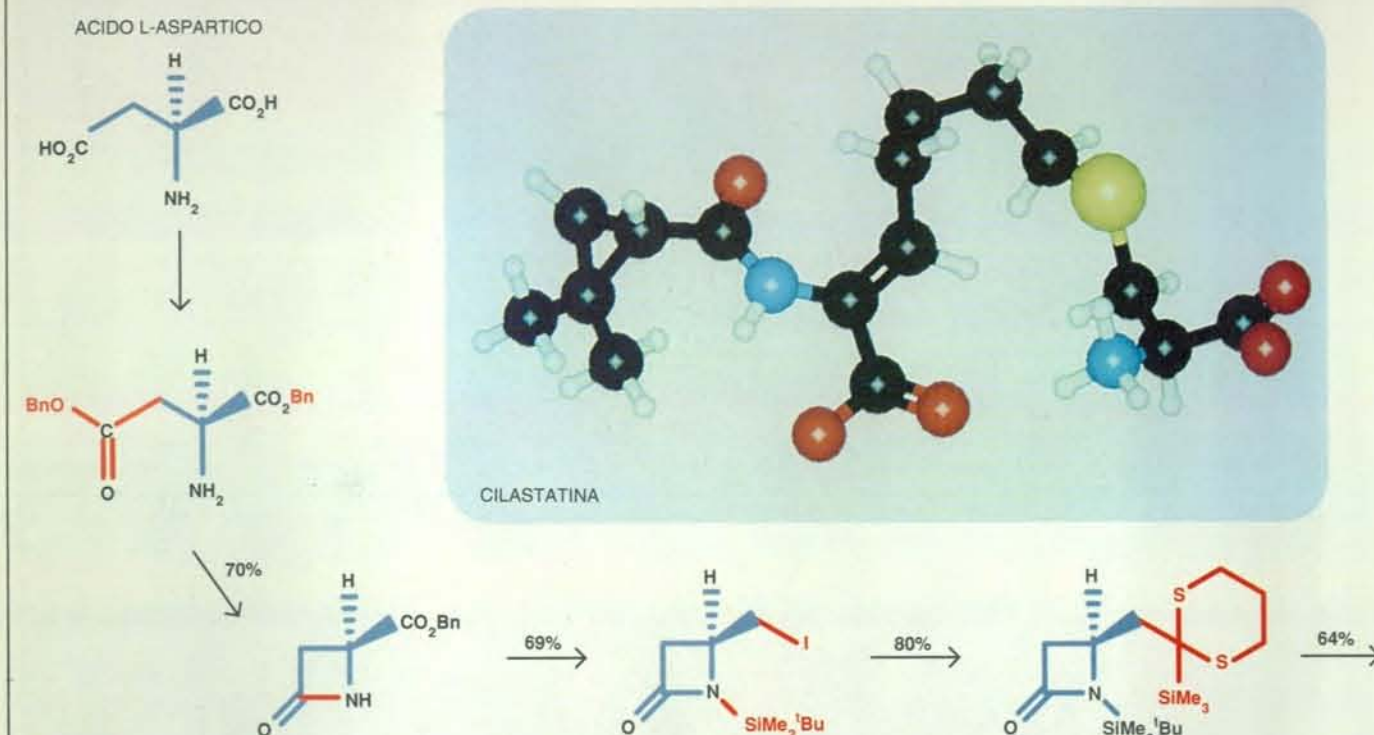
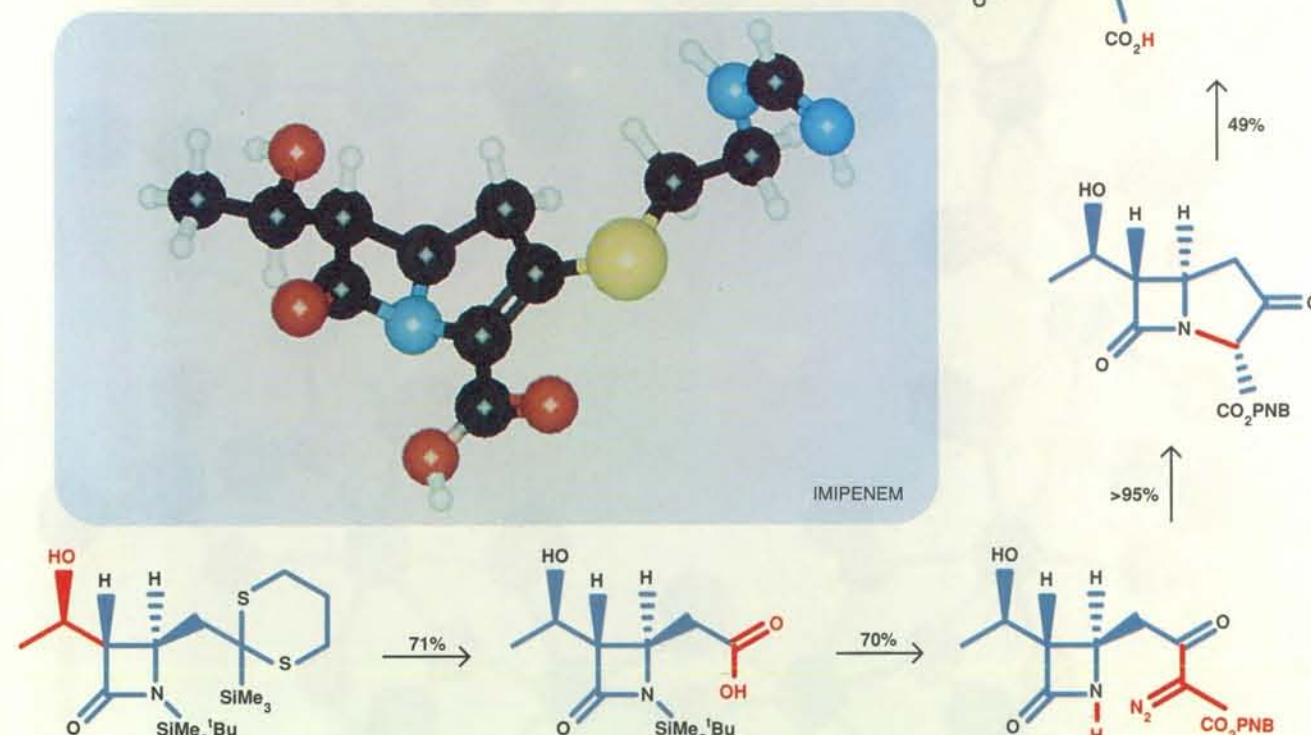
Wöhler mischiò insieme due sostanze, le riscaldò e ottenne un risultato inaspettato. Dal 1828 a oggi sono successe molte cose. Per spiegare che cosa vuol dire oggi fabbricare molecole e come i processi naturali si intreccino con quelli artificiali in questa attività creativa, permettetemi di raccontare la storia della sintesi di due sostanze: l'imipenem-cilastatina e la *ferric wheel* o «ruota ferrica» [così chiamata, fantasiosamente, per assonanza con la *ferris wheel*, la ruota panoramica del luna park].

La imipenem-cilastatina è uno dei più potenti antibiotici sul mercato, una fonte primaria di introiti per la Merck & Co. Questo farmaco non è formato da una sola molecola, ma da due sostanze presenti in precisa proporzione, imipenem e cilastatina. Imipenem e cilastatina sono nomi gergali; i corrispettivi nomi sistematici sono un po' più complicati; per esempio l'imipenem è acido [5R-[5α, 6α(R*)]]-6-(1-idrossietil)-3-[[2-[(imminometil)ammino]etil]tio]-7-osso-1-azabiciclo[3.2.0]epta-2-ene-2-carbossilico.

La imipenem-cilastatina fu creata gra-

Questo particolare dell'affresco di Raffaello *La scuola di Atene* raffigura Platone e Aristotele e, in maniera simbolica, i rispettivi approcci alla conoscenza. La mano di Aristotele è rivolta verso terra; Platone invece indica il cielo. Aristotele cercò le proprie risposte nella natura; Platone andò in cerca dell'idea. I chimici dovrebbero emulare Platone o Aristotele?

La sintesi di un antibiotico

La imipenem-cilastatina è un antibiotico formato da due sostanze, imipenem e cilastatina. L'imipenem è una forma leggermente modificata della tienamicina, che in natura viene prodotta da una muffa. I chimici hanno sviluppato un procedimento (di cui vengono mostrate le fasi principali) per produrre la tienamicina in maniera più efficiente rispetto ai processi naturali di cui si era a conoscenza. Le formule schematiche sono una tipica notazione chimica; in esse non sono indicati tutti gli atomi. I vertici privi di simbolo atomico corrispondono ad atomi di carbonio. La maggior parte degli atomi di idrogeno è stata omessa. È possibile dedurre la posizione degli idrogeni mancanti tenendo presente che ogni atomo di carbonio deve formare quattro legami. Una freccia indica ogni trasformazione chimica nel corso del processo, mentre il valore percentuale sopra le frecce equivale alla resa sperimentale. I simboli raffigurati in rosso indicano cambiamenti della struttura. Nei modelli della cilastatina e dell'imipenem, in nero viene indicato il carbonio, in azzurro l'azoto, in giallo lo zolfo, in rosso l'ossigeno e in bianco l'idrogeno. I simboli a cuneo, pieni o tratteggiati, indicano altri dettagli della configurazione della molecola, e segnatamente quegli atomi che si trovano al di sopra o al di sotto del piano molecolare.

zie a un espediente, che simula gli espedienti naturali dell'evoluzione. L'imipenem è di per sé un buon antibiotico, ma viene rapidamente degradato da un enzima renale, e ciò esclude la possibilità di usare il farmaco per le infezioni del tratto urinario. I ricercatori della Merck trovarono nella loro raccolta di campioni una sostanza interessante, sintetizzata negli anni quaranta e capace di inibire l'attività di quel cocciuto enzima. La modificarono per renderla più attiva e ottennero la cilastatina. Naturalmente si sperimentarono gli effetti della combinazione antibiotico-inibitore dell'enzima e la miscela si rivelò efficace. L'imipenem deriva da un prodotto naturale; la cilastatina no. Entrambi vengono però sintetizzati artificialmente nel processo industriale. Ritornerò su questo punto dopo aver approfondito la storia di uno dei due componenti, l'imipenem.

Esso fu preparato negli anni settanta dal gruppo di Burton G. Christensen. Si tratta della forma leggermente modificata di un altro antibiotico, la tienamicina, la quale, a sua volta, fu scoperta analizzando campioni di terreno del New Jersey. È prodotta da una muffa, *Streptomyces cattleya*, così chiamata perché il suo color lavanda rammenta quello delle orchidee del genere *Cattleya*. Questa muffa è una vera fabbrica di medicinali: oltre alla tienamicina produce molti altri tipi di antibiotici.

Purtroppo la tienamicina non è chimicamente stabile ad alte concentrazioni e il microrganismo che la produce non è particolarmente generoso: i comuni processi di fermentazione, perfezionati dall'industria farmaceutica nel corso degli ultimi 50 anni, non arrivavano a produrre una quantità sufficiente di questa molecola. Così si cercò di ottenere maggiori quantità di tienamicina in laboratorio.

La produzione di tienamicina richiedeva 21 passaggi principali, ciascuno dei quali implicava una serie di operazioni, come dissoluzione, riscaldamento, filtrazione, cristallizzazione. Tra la sostanza di partenza - un comune amminoacido, l'acido L-aspartico - e il prodotto desiderato - la tienamicina - venivano isolate e purificate 20 altre molecole. Di queste, solo otto sono indicate nello schema di reazione mostrato in queste due pagine.

La prima impressione che si riceve è che si tratti di una procedura molto complessa. Ma la complicazione è inevitabile: è il contraltare di laboratorio alla complessità biochimica nostra e dei batteri. Ci piacerebbe che esistessero «proiettili magici» di estrema semplicità, ma il mondo reale è bello e complicato, e dovremmo cercare di dialogare con questa ricchezza.

Per avere un'idea della difficoltà del processo, almeno in termini di sudore, se non di lacrime e sangue, dobbiamo esaminare la parte sperimentale dell'articolo che riguarda la sintesi della tienamicina. Riportiamo un estratto di quel protocollo sperimentale, in cui viene descritto un passaggio critico della sintesi, la trasformazione dal composto 8 al composto 9:

> Una sospensione di diazochetoestere 8 (3,98 g, 10,58 mmol) e di acetato di rodio(II) in forma di dimero (0,04 g, 0,09 mmol) in toluene anidro (250 ml) venne purificata a fondo con azoto, e poi fu riscaldata, mescolando, in un bagno di olio mantenuto a 80 gradi Celsius. Dopo averla riscaldata per due ore, si tolse la miscela di reazione dal bagno e la si filtrò, finché era ancora tiepida, facendola passare attraverso uno strato di solfato di magnesio anidro. Il filtrato fu fatto evaporare sotto vuoto in modo da ottenere chetoestere biciclico 9 (3,27 g, 89%) sotto forma di solido biancastro...

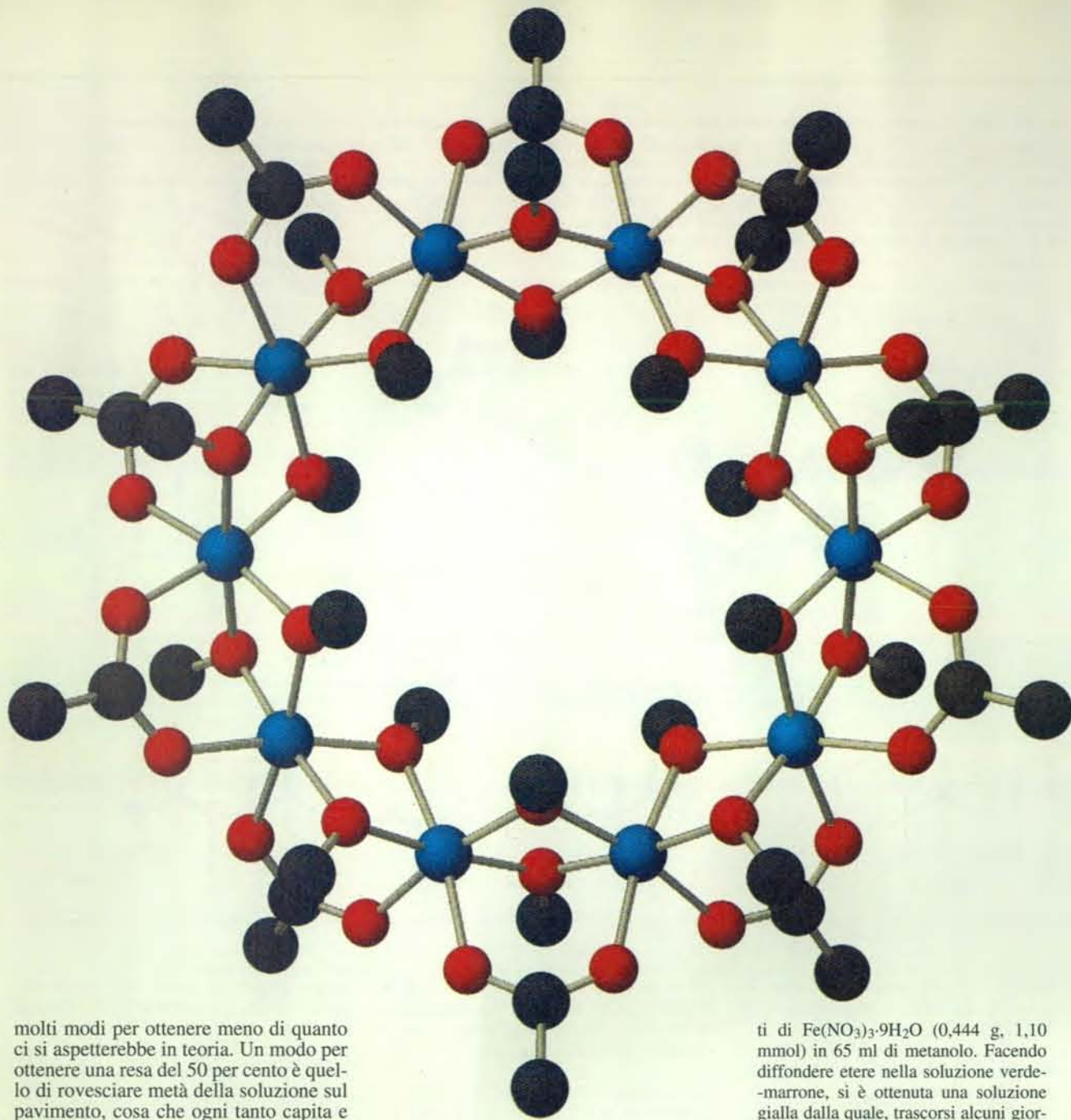
Di sicuro questo resoconto gergale e troppo lineare di una procedura sperimentale è idealizzato; questo è il modo pulito, ottimizzato in cui le cose sono state descritte *alla fine*. Non è certamente il modo in cui si sono verificate la prima volta. A parte questa considerazione, ci si rende conto che il lavoro equivale a una sequenza di operazioni che richiedono tempo e fatica. A volte, proprio come ci capita di immaginare, di fronte a una poesia, che le parole siano sgorgate spontanee da una mente ispirata, ci dimentichiamo di quanto sia faticoso l'atto creativo. Perfino il Creatore si riposò il settimo giorno.

È interessante vedere come cambiano le procedure sperimentali quando lo stesso processo viene svolto a grande scala. Non si può fabbricare una quantità di tienamicina del valore di centinaia di milioni di dollari nello stesso modo in cui se ne producono pochi grammi in laboratorio. Qui di seguito è riportata la descrizione dello stesso passaggio per la sintesi industriale:

> I solidi contenenti 200 kg di 8 vengono introdotti in 476 galloni di $MeCl_2$ all'interno del serbatoio TA-1432. Nel frattempo il reattore ST-1510 è pulito con 200 galloni di $MeCl_2$ bollente. Il materiale semiliquido viene trasferito in ST-1510, dopodiché si lavano i condotti con 50 galloni di $MeCl_2$. Altri 400 galloni di $MeCl_2$ anidro sono aggiunti a ST-1510 e alla camicia del reattore si applica acqua calda (65 gradi Celsius) per concentrare la miscela a 545 galloni; a questo punto la concentrazione della fanghiglia KF (Karl Fischer) è di circa 0,5 grammi per litro di H_2O. I distillati vengono condensati e raccolti in un altro serbatoio.

Preparare filetto alla Stroganoff per mille persone non è la stessa cosa che cucinarlo a casa per quattro.

La sintesi della tienamicina equivale a una costruzione in cui si parte da pezzi semplici e si arriva a un risultato complesso. Ha molti aspetti in comune con l'architettura. Una struttura che serve in una fase intermedia della costruzione può essere molto più complicata sia della struttura iniziale sia di quella finale: si pensi alle impalcature. La sintesi chimica è una sfida locale all'entropia, proprio come lo sono gli edifici e le città. Il paragone con l'architettura è così stretto che può essere fuorviante, facendoci dimenticare che la costruzione chimica è molto diversa e che in ciò sta la sua bellezza. In un matraccio possono esserci 10^{23} molecole che si muovono velocemente e spesso si scontrano. Senza alcun intervento esterno, rispettando solo le ferree regole della termodinamica, queste molecole rimescolano gli elettroni, fanno e sciolgono legami ed eseguono i nostri ordini. Se siamo fortunati, il 99 per cento di esse si comporta in questo modo.

I chimici possono facilmente calcolare quanto prodotto si dovrebbe ottenere data una certa quantità di materiale di partenza. Questa è la cosiddetta resa teorica. La quantità che si ottiene di fatto è la resa sperimentale. Non è possibile ottenere qualcosa dal nulla, ma ci sono

molti modi per ottenere meno di quanto ci si aspetterebbe in teoria. Un modo per ottenere una resa del 50 per cento è quello di rovesciare metà della soluzione sul pavimento, cosa che ogni tanto capita e che quindi non stupisce particolarmente. Ma anche se si esegue ogni passaggio nel modo più pulito possibile, si può non ottenere il risultato desiderato, bensì trasformare il 70 per cento della reazione in una schifezza. Anche questo non stupisce, poiché conferma che non sempre la mente riesce a controllare la materia. Le rese sperimentali sono un parametro non solo dell'efficienza, essenziale da un punto di vista industriale, ma anche dell'eleganza e del controllo.

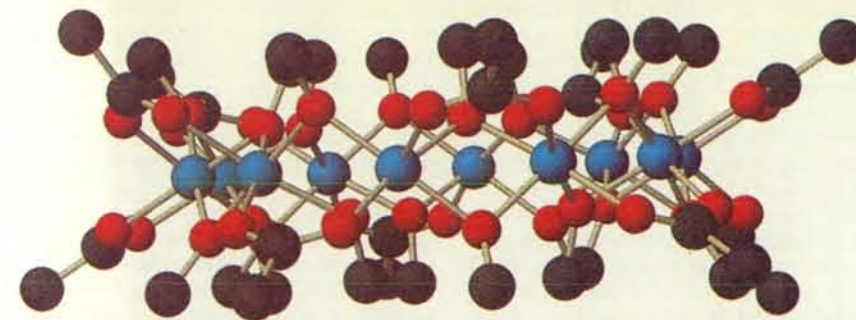
Ci sarebbero molte altre cose da raccontare sulle sintesi organiche programmate. Ma ora mi devo occupare del mio secondo esempio: la ruota ferrica.

Stephen J. Lippard e Kingsley L. Taft del Massachusetts Institute of Technology sintetizzarono la ruota ferrica, $[Fe(OCH_3)_2(O_2CCH_2Cl)]_{10}$, mentre studiavano modelli molecolari delle reazioni inorganiche che si verificano nei sistemi biologici. Per esempio, un aggregato di atomi di ferro e di ossigeno fa parte del nucleo di importanti proteine, come la emeritrina, la reduttasi ribonucleotidica, la metano monossigenasi e la ferritina (questi termini non sono d'uso comune, ma si riferiscono a molecole indispensabili per la vita).

Nel corso della loro approfondita ricerca su queste sostanze, Lippard e Taft eseguirono una reazione estremamente semplice. La sua semplicità risalta dalla lettura della parte sperimentale del lavoro, qui riportata per intero:

> Il composto 1 è stato preparato facendo in modo che il monocloroacetato analogo dell'acetato basico di ferro, $[Fe_3O(O_2CCH_2Cl)_6(H_2O)_3](NO_3)$ (0,315 g, 0,366 mmol), reagisse con 3 equivalenti di $Fe(NO_3)_3 \cdot 9H_2O$ (0,444 g, 1,10 mmol) in 65 ml di metanolo. Facendo diffondere etere nella soluzione verde-marrone, si è ottenuta una soluzione gialla dalla quale, trascorsi alcuni giorni, si sono depositati sia cristalli marrone dorato del composto 1 sia un precipitato giallo.

Analizzando con il metodo della diffrazione di raggi X i cristalli marrone dorato, Lippard e Taft determinarono la disposizione degli atomi nella molecola. La struttura consiste di 10 ioni ferrici (cioè ferro allo stato di ossidazione 3) disposti grosso modo in cerchio. Ogni atomo di ferro è unito ai vicini tramite ponti metossido e carbossilato, formando quella che gli scopritori definiscono una «ruota ferrica molecolare».

Nessuno può negare la bellezza di questa molecola. Certo non arriva a fruttare mezzo miliardo di dollari all'anno come la imipenem-cilastatina; anzi, probabilmente la sua produzione costa ai contribuenti statunitensi diverse migliaia di dollari. Ma non conosco alcun chimico che non ammiri questa elegante creazione. Forse un giorno la ruota ferrica troverà un'applicazione, forse contribuirà a spiegare la funzione delle proteine che contengono il ferro. Non è molto importante: a me questa molecola procura un godimento spirituale paragonabile a quello che provo ascoltando un trio di Haydn che amo particolarmente.

La ruota ferrica, la cui formula è $[Fe(OCH_3)_2(O_2CCH_2Cl)]_{10}$, è un esempio di struttura ideale. La molecola è formata da atomi di ferro (*in blu*), ossigeno (*in rosso*) e carbonio (*in nero*). Per evidenziare la simmetria, sono stati omessi cloro e idrogeno.

Perché questa molecola è bella? Perché la sua simmetria tocca direttamente l'anima. Fa vibrare la nota di un'idea platonica. Forse dovrei paragonarla alla canzone *Amazing Grace* cantata da Judy Collins, anziché al trio di Haydn. Le linee melodiche del trio sono davvero notevoli, ma la composizione colpisce grazie al contrappunto, lo strumento della complessità. La ruota ferrica è invece paragonabile a una pura melodia.

Se dovessimo scrivere la reazione di sintesi della ruota ferrica, basterebbe una sola freccia, dal cloroacetato di ferro e dal nitrato ferrico al prodotto. È un tipo di sintesi molto particolare: in pratica il prodotto si autoassembla per raggiungere lo stato di gloria finale. Quando osservo un tale processo, molto più tipico dei sistemi inorganici che non di quelli organici, mi chiedo immediatamente che cosa sto perdendo. Il chimico svedese Sture Forsén ha espresso molto bene la frustrazione che si prova quando non si è in grado di osservare gli stadi intermedi di una reazione:

> Il disagio dello scienziato è stato paragonato a quello di uno spettatore che assista a una versione di una tragedia classica drasticamente ridotta. Si prenda per esempio l'*Amleto* di Shakespeare e si immagini che allo spettatore siano mostrate solo le scene iniziali del primo atto e la scena finale dell'ultimo atto. Vengono introdotti i personaggi principali, poi cala il sipario per il cambio della scena e, quando si rialza, vediamo molti «cadaveri» e pochi sopravvissuti. Per una persona ignara della trama non è certo facile ricostruire che cosa sia successo in mezzo.

Gli anelli, siano essi ferrici o ferrosi, non si autoassemblano di punto in bianco. Spetta a noi il compito di arrivare a capire in che modo questi ponti e questi atomi di ferro si uniscano tra di loro.

Alcuni chimici non vedono di buon occhio una reazione di autoassemblaggio che si svolge in un solo stadio; in genere costoro preferiscono le sintesi programmate a più stadi, del tipo di quella della tienamicina, che richiedono un notevole sforzo mentale e sono intellettualmente stimolanti. Le sintesi a uno stadio sono particolarmente comuni nella chimica dello stato solido, per produrre materiali che si estendono all'infinito in una, due o tre dimensioni. I superconduttori ad alta temperatura sono un buon esempio di molecole sintetizzate in questo modo. La loro sintesi non mostra alcun controllo della mente sulla materia: sembra avvenire per magia.

Ho esagerato, ma questo è un modo di pensare diffuso all'interno della nostra comunità. Se potessi mettere con le spalle al muro un derisore da due soldi dell'autoassemblaggio, in genere un chimico organico, e lo potessi coinvolgere in un dialogo socratico, inizierei chiedendogli: «Quando hai fabbricato un diamante per me, ultimamente?» Il diamante è una struttura tridimensionale bella e semplice (e naturale!). Contiene anelli a sei atomi, che sono la base della chimica organica. Gli anelli a sei atomi di carbonio sono facili da ottenere in una molecola isolata, ma il diamante in quanto tale può essere preparato solo con tecniche che i chimici organici considerano poco ortodosse, per esempio per mezzo di scariche elettriche che formano un plasma nel metano o sottoponendo la grafite a pressione elevata.

I chimici organici sono maestri nell'esercitare controllo in zero dimensioni. A un pezzo di carbonio, magari asimmetrico, aggiungono un altro pezzo; piano piano e con fatica viene costruita una struttura complessa. (La tienamicina è una molecola abbastanza semplice in confronto a ciò che si può ottenere oggigiorno.) Una sottobranca di chimici organici ha imparato a esercitare il controllo anche in una dimensione: costoro, i chimici dei polimeri, costruiscono i loro prodotti sotto forma di catene. (Di solito non ricevono i riconoscimenti che meriterebbero, ma in compenso guadagnano un bel po' di quattrini.)

Ma in due o in tre dimensioni, non si riesce a ottenere alcuna sintesi. Manca quasi del tutto una metodologia che permetta di formare su richiesta strutture di notevoli dimensioni, instabili ma persistenti. Si può anche considerare la cosa in maniera positiva e pensare che questo sarà un campo in cui la chimica del futuro farà sicuramente molti progressi.

Le sintesi, come gli esseri umani, non si prestano a generalizzazioni. Ciascuna è diversa dall'altra; ognuna ha i propri pregi e i propri difetti e può insegnarci qualcosa.

Nell'arte della sintesi si nascondono due paradossi. Il primo è che l'atto della sintesi è esplicitamente umano e quindi innaturale, anche se ci si propone di ottenere un prodotto già esistente in natura. Il secondo è che nella sintesi di molecole ideali, laddove sembrerebbe di dover operare in modo prettamente «innaturale», bisogna invece talvolta cedere alla natura. Cercherò di spiegare che cosa intendo rifacendomi alle due sintesi di cui ho appena parlato.

L'imipenem si ottiene dalla tienamicina. Questa è una sostanza naturale, ma, in base a calcoli di convenienza sia di tipo economico sia di tipo chimico, si decise che ai fini della produzione commerciale fosse meglio ricavare la tienamicina per via sintetica.

In questo caso la molecola naturale fu sicuramente usata dai chimici come modello a cui ispirarsi, ma ovviamente essi non fabbricarono la tienamicina in laboratorio secondo le stesse modalità con cui viene prodotta dalla muffa. Quest'ultima ha le sue complicate fabbriche chimiche: gli enzimi, che sono stati modellati dall'evoluzione. Solo di recente abbiamo imparato in che modo utilizzare l'ingegneria genetica per sfruttare a nostro vantaggio quelle fabbriche, quando non addirittura gli interi organismi.

Siamo diventati esperti nelle tecniche chimiche di laboratorio, che sono più semplici di quelle evolute dagli organismi biologici. In nessun modo Christensen e il suo gruppo avrebbero potuto imitare nei dettagli un enzima della muffa. Essi sapevano invece di poter imitare, solo in piccola parte e secondo modalità del tutto diverse, l'operato di quell'umile muffa, allo scopo di ottenere la tienamicina. Il loro obiettivo era un prodotto naturale, il processo che usavano per ottenerlo no.

Per fabbricare la tienamicina, Christensen e collaboratori usarono un gran numero di reagenti, di tipo sia naturale sia sintetico. Per esempio una delle loro trasformazioni - la sintesi del composto 3 (*si veda la finestra alle pagine 74-75*) - utilizza un composto del magnesio, $(CH_3)_3CMgCl$, noto anche come reagente di Grignard. I composti di magnesio sono abbondanti in natura (come

testimoniano il sale di Epsom e la clorofilla), ma il reagente in questione, che ora viene usato comunemente nella sintesi chimica, fu preparato da Victor Grignard verso la fine del secolo scorso. La creazione del composto 3 richiede anche un trattamento sia con acido cloridrico sia con cloruro d'ammonio, che sono entrambi prodotti naturali. (Nello stomaco vi è una concentrazione di acido cloridrico di poco inferiore a quella usata in questa reazione e il cloruro d'ammonio corrisponde al «sale ammoniaco» degli alchimisti.) Ma anche se queste molecole sono disponibili in natura, è molto più facile ottenerle in un impianto chimico.

Dal momento che in definitiva tutto deriva dalla terra, dall'aria o dall'acqua, qualsiasi reagente non naturale usato nella sintesi alla lunga deriva da precursori naturali organici o inorganici. Il materiale di partenza nella sintesi dell'imipenem è un amminoacido, l'acido aspartico.

Ora si prenda in considerazione la ruota ferrica, più bella e «innaturale». Essa fu ottenuta semplicemente facendo reagire in metanolo (un solvente naturale) due molecole sintetiche, il monocloroacetato di ferro e il nitrato ferrico. Probabilmente il metanolo fu prodotto per via sintetica; i due reagenti che contengono il ferro derivano da reazioni del ferro metallico, che a sua volta viene estratto dai minerali ferrosi. E il trucco finale è il modo in cui il tutto viene assemblato; i pezzi della molecola sembrano cadere magicamente al proprio posto (autoassemblaggio). Che cosa potrebbe essere più naturale del lasciare che le cose accadano spontaneamente, rispettando le ferree regole dell'entropia?

È chiaro che nella fabbricazione sintetica di una molecola naturale (la tienamicina) o di una artificiale (la ruota ferrica) vengono usati reagenti e solventi sia di tipo naturale sia di tipo sintetico in un delicato equilibrio tra il lasciare che le cose seguano il proprio corso e l'aiutarle a procedere. Si può dire che la sola costante sia il cambiamento, la trasformazione.

Possiamo porci altre domande sulla psicologia della creazione chimica. Per quali molecole val la pena di spendere le nostre energie creative? Non è forse decisamente meglio cercare di ottenere qualcosa di completamente nuovo?

Negli ultimi 40 anni quattro bei poliedri del carbonio hanno catturato l'interesse dei chimici organici che si occupano di sintesi; il tetraedrano (C_4H_4), il cubano (C_8H_8), il dodecaedrano ($C_{20}H_{20}$) e il buckminsterfullerene (C_{60}). Il cubano è alquanto instabile a causa della tensione cui è sottoposto ogni atomo di carbonio. (Nel cubano l'angolo fra tre atomi qualsiasi di carbonio è 90 gradi, ma il carbonio «preferisce» formare un angolo di 109,5 gradi con i suoi vicini.) Il C_{60} è a sua volta un po' distorto, perché non è planare e contiene anelli di 5 atomi. Il tetraedrano è particolarmente instabile. Per osservarlo, bisogna creare opportune condizioni di temperatura e di solvente; e anche in quel caso, non si forma la molecola progenitrice, ma solo un suo «derivato sostitutivo», in cui al posto dell'idrogeno c'è un voluminoso gruppo organico.

Per quanto ne sappiamo, il tetraedrano, il cubano e il dodecaedrano non si trovano sulla Terra allo stato naturale. Il C_{60} è stato individuato nella comunissima fuliggine e in una antica roccia ricca di carbonio, la shungite. Può darsi che lo si trovi anche da qualche altra parte. Comunque, tutte e quattro le molecole sono state considerate obiettivi di sintesi almeno 20, e talvolta addirittura 50 anni prima che si riuscisse a ottenerle. Alcuni dei migliori chimici di tutto il mondo tentarono senza successo di sintetizzarle. Le sintesi del cubano e soprattutto del dodecaedrano furono acquisizioni «monumentali» nella chimica dei prodotti sintetici.

Il caso del C_{60} fu diverso. La bella forma poliedrica fu segnalata per la prima volta da alcuni chimici teorici. Sulla base dei loro calcoli essi pensavano che la struttura dovesse essere stabile; talvolta però le loro indicazioni si erano rivelate inattendibili ed è per questo che la segnalazione fu ignorata dai chimici sperimentali e da altri teorici. Qualche volta è difficile notare le spalle dei giganti su cui poggiamo, soprattutto se siamo intenti a guardare in avanti. Anch'io ho proposto una modificazione metallica del carbonio, diversa dalla grafite e dal diamante, che non è ancora stata sintetizzata, e anche se sono indubbiamente più visibile nella comunità dei chimici di quanto non siano coloro che hanno segnalato il buckminsterfullerene, nessuno, probabilmente per buoni motivi, ha prestato molta attenzione alla mia fantasticheria. Vediamo solo ciò che vogliamo vedere.

Un eccellente chimico organico di mia conoscenza, Orville L. Chapman dell'Università della California a Los Angeles, ha pensato indipendentemente alla struttura del C_{60} e ha dedicato molto tempo alla sua preparazione pianificata e sistematica. Dopo tutto, questa era una molecola «semplice»; non era certo estesa come il reticolo ripetitivo degli atomi di carbonio che formano un diamante. Quindi avrebbe dovuto essere possibile sintetizzarla senza troppi problemi. Nonostante i molti sforzi compiuti in un arco di tempo di 10 anni, Chapman e i suoi studenti non riuscirono a raggiungere il loro obiettivo.

La prima prova, indiretta ma decisiva, dell'esistenza del C_{60} fu ottenuta in un ramo molto diverso della nostra scienza, la chimica fisica. Il merito della scoperta va spartito tra Richard E. Smalley e Robert F. Curl della Rice University e Harold W. Kroto dell'Università del Sussex. Questi scienziati riuscirono a dimostrare la presenza di minuscole quantità di C_{60} in fase gassosa, assegnarono un nome alla molecola e, cosa più importante, ne dedussero la struttura. Si può dire che riuscirono a produrlo? Io penso di sì. Non importa molto né a me né a coloro che credono nel loro esperimento che essi abbiano ottenuto «solo» 10^{10} molecole anziché le 10^{20} di cui abbiamo bisogno per formare un minuscolo cristallo. Ma alcuni chimici organici, e credo che fossero molti, non ci credettero. Volevano vedere la sostanza.

Grammi di buckminsterfullerene furono ottenuti grazie alla sintesi messa a punto da Donald R. Huffman dell'Università dell'Arizona e Wolfgang Krätschmer e Konstantinos Fostiropoulos del Max-Planck-Institut für Kernphysik di Heidelberg. Si può dire che vaporizzare il carbonio con un arco elettrico in un'atmosfera di elio (esattamente ciò che essi fecero) è più o meno altrettanto poco ortodosso quanto dirigere un laser sulla grafite (come avveniva nella sintesi di Smalley-Kroto-Curl). Ma sicuramente con questo metodo si ottiene molto C_{60}, in quantità sufficiente per determinarne la struttura con i metodi tipici della chimica organica, e per convincere tutti i chimici della sua forma a «pallone da calcio» (si veda l'articolo *I fullereni* di Robert F. Curl e di Richard E. Smalley in «Le Scienze» n. 280, dicembre 1991).

Penso che molti chimici avrebbero preferito che il C_{60} fosse stato preparato con una sintesi pianificata. Mi fa piacere invece - proprio per rendere il mondo un po' meno razionale di quanto non vorremmo che fosse - che sia stato ottenuto in maniera casuale e fortuita. [Gli anglosassoni impiegano, per rendere l'idea di «scoperta inattesa e fortuita», il termine *serendipity*, coniato da Horace Walpole.]

Ebbene, sia che si tratti di un chimico in laboratorio o di un maestro vasaio giapponese che depone accuratamente materia organica intorno agli oggetti di ceramica nel suo forno Bizen, la sorte favorisce la mente ben disposta. Bisogna conoscere (per alcuni si tratta di intuizione) il modo in cui si devono variare le condizioni della produzione dell'arco elettrico o la disposizione delle foglie nel forno. Bisogna avere gli strumenti e l'intuizione per dedurre una struttura dalle poche righe confuse di uno spettro e per respingere indizi fuorvianti. E bisogna avere il coraggio di buttare via un vaso che non è riuscito bene e di imparare da un primo tentativo come correggere i successivi.

Molte sintesi chimiche, anche se fanno parte di un grande disegno complessivo, procedono per stadi che sono caratterizzati da *serendipity*. Si vuole formare un legame in questo punto, ma non ci si riesce. Così si asseconda un presentimento, seguendo tutto tranne che un metodo scientifico codificato. Si sa che se una reazione funziona, potrà poi essere razionalizzata, cioè si potrà costruire un'argomentazione su misura sufficientemente pulita da impressionare i propri colleghi. Alla fine si può fare in modo che la maledetta reazione funzioni, se essa rappresenta un passaggio ineliminabile del progetto.

Dal momento che il caso contribuisce a sabotare ogni progetto, è molto probabile che nel corso di ogni sintesi programmata si incontri un passaggio che non funziona in nessun modo noto. Così si inventerà un nuovo modo, che si andrà ad aggiungere al repertorio dei chimici e che aiuterà altri scienziati in tutto il mondo a risolvere lo stesso problema. Alcuni chimici che si occupano di sintesi - per esempio E. J. Corey della Harvard University, un grande maestro dell'arte - hanno un talento speciale non solo nel produrre molecole interessanti, ma anche nello sfruttare l'opportunità offerta dalle sintesi per elaborare una metodologia brillante e inedita, che sia applicabile ad altre sintesi.

Quando si programma una sintesi, indipendentemente dal fatto che abbia come obiettivo una molecola naturale o artificiale, si cerca di sopprimere la natura aleatoria dell'impresa. Si vuole dare un'immagine di controllo totale da parte della mente sulla materia. Quando la molecola prodotta è del tutto imprevista, come nell'esempio della ruota ferrica, diventa molto difficile nascondere il ruolo giocato dal caso. Quest'ultimo però svolge un ruolo in tutte le sintesi, un ruolo non riconosciuto che vivacizza la situazione.

Ma torniamo ora alla natura e al nostro tentativo di emularla. O di superarla. Possiamo produrre sostanze dotate di proprietà superiori rispetto a quelle realizzate dalla natura? Rispondo di sì, pur riconoscendo che l'affermazione «superiore alla natura» è alquanto pesante e antropocentrica e che dovrebbe immediatamente evocare preoccupazioni di tipo ecologico.

Nelle reti da pesca si usa il nylon al posto del cotone, nelle calze da donna si usa il nylon al posto della seta. Nessuno, men che meno i pescatori del Terzo Mondo, ritornerebbe alle vecchie reti. Forse qualcuno tornerebbe alle calze di seta, ma si tratterebbe solo del capriccio di qualche ricco, un caso isolato non particolarmente significativo. Esistono nuovi materiali chimici e nuove combinazioni di vecchi materiali per la ricostruzione dei denti. Questi materiali hanno cambiato la vita degli anziani e i vantaggi che procurano non possono essere sottovalutati.

Tuttavia il pensiero che noi uomini possiamo fare meglio della natura è provocatoriamente arrogante. Dal momento che abbiamo tentato di migliorare la natura (mentre non siamo riusciti a controllare la cosa più naturale di tutte, la nostra spinta a procreare), abbiamo introdotto tante e tali trasformazioni che abbiamo contaminato il nostro stesso nido e ci siamo intromessi nei grandi cicli del pianeta. Dobbiamo affrontare la realtà e ammettere che l'evoluzione naturale procede troppo lentamente per tenere il passo dei nostri cambiamenti. Questa è una preoccupazione che, esattamente come il criterio dell'utilità, dovrebbe guidare le sintesi chimiche su scala industriale del futuro.

Voglio accennare a un'altra forma di arroganza umana che è implicita nel «dramma» intellettuale della sintesi chimica. Un chimico francese, Alain Sevin, è riuscito a esprimerla bene:

> L'incredibile ricchezza e fantasia della Natura è un atto di sfida per l'Uomo, che si sente così impegnato a superarla in ogni campo: volare più veloce di un uccello e tuffarsi più in profondità delle balene... Stiamo interpretando Prometeo in una recita senza fine che ora è giunta all'atto molecolare.

Siamo irresistibilmente portati a trasformare. Abbiamo imparato a farlo molto bene. Ma questa «rappresentazione teatrale» non è una commedia.

Se le sintesi chimiche potessero essere simboleggiate da una sola immagine, la mano protesa di Prometeo che porta il fuoco all'umanità potrebbe essere

DODECAEDRANO

BUCKMINSTERFULLERENE

Quattro poliedri basati sul carbonio vennero individuati diverse decine di anni fa come possibili obiettivi di sintesi chimica. Il buckminsterfullerene fu prodotto nel 1985 e in seguito ci si accorse che era possibile trovarlo in natura. Il tetraedrano, la più semplice di queste strutture, non è ancora stato sintetizzato. Il tetraedrano, il cubano e il dodecaedrano sono costituiti di atomi di carbonio (*in nero*) e idrogeno (*in rosso*), mentre il buckminsterfullerene è formato unicamente da carbonio.

adatta allo scopo. Prometeo, un nome che significa «preveggenza», simboleggia quell'elemento del progetto che equivale all'approfittare proficuamente della creazione casuale. Il fuoco è appropriato perché fa avanzare la trasformazione. La mano di Prometeo è il simbolo della creazione, come la mano di Dio che tocca Adamo nell'affresco di Michelangelo, le mani che si agitano nella discussione polemica del *Cristo tra i dottori* di Dürer, l'infinita varietà di mani che sono state scolpite da Rodin. Le mani benedicono, accarezzano e nascondono, ma più di tutto modellano.

L'arte dello scultore ha delle analogie con il complesso modo di muoversi di un chimico, in equilibrio tra ciò che è naturale e ciò che non lo è. Rodin, per compiere un atto creativo di tipo umano, fa un abbozzo, poi plasma con le mani (servendosi di utensili) un modello in scala eppure «realistico», ottenendo la scultura di una mano in un materiale artificiale (come il bronzo) ma di origine naturale (rame e stagno). Usa un procedimento (tramite stampi) dalle complesse fasi intermedie. Lo scultore crea qualcosa di estremamente reale, il cui pregio può essere quello di rievocare nelle nostre menti l'idea.

Margaret Drabble ha scritto che Prometeo è «saldamente radicato nel mondo reale della fatica, del pericolo e del dolore». Senza la sintesi chimica, non ci sarebbero l'aspirina, il cortisone, la pillola anticoncezionale, gli anestetici, la dinamite. Le conquiste della sintesi chimica sono strettamente legate al nostro tentativo di rompere le pastoie della malattia e della povertà. Nella ricerca dell'idea, nella fabbricazione di oggetti reali, la mente e le mani trovano la loro occupazione.

BIBLIOGRAFIA

SALZMANN T. N., RATCLIFFE R. W., BOUFFARD F. A. e CHRISTENSEN B. G., *A Stereocontrolled, Enantiomerically Specific Total Synthesis of Thienamycin* in «Philosophical Transactions of the Royal Society of London», serie B, 289, n. 1036, 16 maggio 1980.

COREY E. J. e XUE-MIN CHENG, *The Logic of Chemical Synthesis*, John Wiley and Sons, 1989.

TAFT KINGSLEY L. e LIPPARD STEPHEN J., *Synthesis and Structure of* $[Fe(OMe)_2(O_2CCH_2Cl)]_{10}$, *A Molecular Ferric Wheel* in «Journal of the American Chemical Society», 112, pp. 9629-9630, 19 dicembre 1990.

ATKINS PETER W., *Molecole*, Nuovi classici della scienza, Zanichelli, Bologna, 1992.

SILVERMAN RICHARD B., *The Organic Chemistry of Drug Design and Drug Action*, Academic Press, 1992.

L'ANGOLO MATEMATICO

di Ian Stewart

Una storia parzialmente vera

Permettete che mi presenti: Epimenide, mentitore di professione. Beh, non è del tutto vero; in realtà mi chiamo Herman Fenderbender e lavoro per un'agenzia di assicurazioni. Ma i miei amici del Club Paradosso mi chiamano Epimenide, e con loro io mento sempre.

Giovedì scorso pioveva, quindi arrivai al club un po' in ritardo. Socrate e Platone erano appoggiati al banco del bar e vicino a loro stava un tipo piccolo e grassottello.

«Questi è Lukasiewicz, il nuovo socio del club» disse Platone in tono giulivo.

«Desolato di conoscerti» dissi con aria di disgusto. «Il mio nome è Zenone.»

«Intende dire che è felice di conoscerti e che il suo nome è Epimenide» spiegò Socrate. «Epimenide mente sempre.»

«Non è vero» replicai. Estrassi dal portafoglio il biglietto da visita. «Questo non è il mio biglietto da visita» commentai porgendolo al nuovo socio. Lukasiewicz lesse su un lato: *L'affermazione sull'altro lato di questo biglietto è vera.* Girò il biglietto e lesse: *L'affermazione sull'altro lato di questo biglietto è falsa.*

«Socrate ha ragione, però: io dico sempre il falso» mi vantai.

Lukasiewicz mi strinse calorosamente la mano. «È per un terzo falso che ho piacere di conoscerti, e le due facce del tuo biglietto da visita sono entrambe vere a metà.»

«Come, scusa?» dissi.

«Lukasiewicz si interessa di logica

sfumata [*fuzzy logic*]» spiegò Platone.

«Invece di due soli valori di verità, 1 per un enunciato vero e 0 per uno falso - disse Lukasiewicz - io sono pronto a prendere in considerazione mezze verità con valore di verità 0,5 o quasi falsità con valore 0,1... in generale, un numero qualsiasi compreso tra 0 e 1.»

«E perché mai si dovrebbe fare una cosa del genere?» chiesi stupefatto.

Lukasiewicz sorrise. «Supponi che io ti dica che il presidente del club assomiglia a Charlie Chaplin. Ti sembra che sia vero?»

«Naturalmente no!»

«Nemmeno nei piedi?»

«Beh, in effetti...»

«Allora non è nemmeno completamente falso.»

«D'accordo: assomiglia un pochino a Chaplin.»

Lukasiewicz si piegò verso di me. Aveva occhi decisamente penetranti. «Quanto gli assomiglia?»

«Circa per il 15 per cento, direi.»

«Bene. Allora la mia affermazione "Il presidente del club assomiglia a Charlie Chaplin" è vera per il 15 per cento. In logica sfumata ha un valore di verità pari a 0,15.»

«Ma questo vuol dire giocare con le parole! Non ha alcun senso!»

Lukasiewicz mi strinse il braccio. «Oh, sì che ce l'ha: serve a risolvere i paradossi. Per esempio, tu sostieni di essere un mentitore assoluto; ma riflettiamo sulla tua affermazione "sto mentendo", o più semplicemente "questa affermazione è falsa". In logica classica è un paradosso, no? Se l'enunciato è vero, allora è falso; se è falso, allora è vero. Per dirla in altri termini, abbiamo un enunciato P con valore di verità p, che è uguale a 1 o a 0, e P dice che il valore di verità di questo enunciato è $1-p$.»

Il diagramma del «tempo di fuga» (*a sinistra*) e un attrattore logico (*qui sotto*) sono stati creati per analizzare un enunciato autoreferenziale. In logica classica, questi enunciati portano di solito alla formulazione di paradossi. Le due illustrazioni si basano sull'affermazione «la falsità stimata di questo enunciato non è diversa dalla sua verità stimata».

«Scusa, non ho afferrato bene.»

«Dunque. Se P è vero, allora la sua negazione, *non* P, è falsa, e il suo valore di verità è 0. E viceversa. Ora, $1-0=1$ e $1-1=0$; pertanto, se il valore di verità di P è p, allora il valore di verità di *non* P è $1-p$.»

«Sì, vedo.»

«Bene. Il problema, ora, è che "questo enunciato" è P: quindi P sta dicendo che il valore di verità di P è $1-p$. Ed è qui che nasce il paradosso. Se $p=0$, allora P ci dice che $p=1-0=1$. E se $p=1$, allora P ci dice che $p=1-1=0$. In entrambi i casi si ha una contraddizione.»

Gli rivolsi un sorriso condiscendente. «Luke, non hai fatto altro che riformulare un asserto del tutto evidente in un complicato linguaggio algebrico.»

«Può darsi» replicò lui sogghignando. «Ma in logica sfumata c'è una soluzione coerente all'equazione $p=1-p$, ossia $p=0,5$. La tua pretesa di dire sempre il falso è una mezza verità, e tutto funziona benissimo. Come vedi, è proprio la tua affermazione a portarci inevitabilmente alla logica sfumata.» Platone gli diede una gran pacca sulla schiena, mentre Socrate quasi si strozzava dal ridere. Io arrossii, ma dovetti riconoscere che aveva ragione.

«E che cosa ne dici del suo biglietto da visita?» chiese Platone. Lukasiewicz stava per parlare, ma lo bloccai. «Lascia che risponda io. Mi sembra che ci siano due enunciati P e Q, con valori di verità p e q. Inoltre, P dice che Q è vero, e Q dice che P è falso. Quindi le equazioni per trovare i valori di verità, rispettivamente per P e per Q, sono

$$p = q \text{ per } P$$

$$q = 1 - p \text{ per } Q.$$

La cosa è assurda se i valori di verità possono essere solo 0 o 1. Ma c'è una soluzione in logica sfumata: $p=q=0,5$. Ciascun lato del mio biglietto da visita, allora, è una mezza verità e non c'è alcun paradosso.»

«Proprio così» disse Lukasiewicz. «Ma c'è ben di più. Quello di cui abbiamo parlato è solo il punto di partenza di un'intera nuova teoria di logica dinamica, formulata da Gary Mar e Patrick Grim del Dipartimento di filosofia della State University of New York a Stony Brook, che mette in relazione i paradossi semantici con la teoria del caos.»

Questa volta fu Socrate ad apparire sconcertato.

«Oh, svegliati. Sai bene che cos'è il caos: semplice dinamica deterministica che porta a un comportamento irregolare, apparentemente casuale. L'effetto farfalla. Quella roba lì.»

«Ma certo, questo lo so bene» disse Socrate con aria irritata. «Era l'idea della logica dinamica a lasciarmi perplesso. Come fa la logica a essere dinamica?»

Lukasiewicz appariva sorpreso. «Che cos'altro può essere quando si discute di enunciati autoreferenziali? L'enunciato stesso ti obbliga a correggere la stima del suo valore di verità, e il valore corretto deve essere corretto ancora, e poi ancora, e poi ancora... Prendiamo il Paradosso del mentitore, il tuo enunciato P: "questo enunciato è falso". Prima ho scritto un'equazione per il suo valore di verità: $p=1-p$. Ma in realtà avrei dovuto scrivere un processo che porti a una continua revisione della stima del valore di verità, ossia $p \leftarrow 1-p$. Se si ammette che P abbia un certo valore di verità p, allora P stesso ci dice di sostituire quel valore con $1-p$. Per esempio, se all'inizio ritieni che P sia vero al 30 per cento, in modo che $p=0,3$, la regola di correzione implica $p=0,7$, che a sua volta implica di nuovo $p=0,3$... e si ottiene una successione infinita di valori di verità che oscillano tra i due valori 0,3 e 0,7. Il paradosso classico, con $p=0$ o $p=1$, porta alla successione 0, 1, 0, 1... che riproduce esattamente il ragionamento logico secondo cui se P è falso, allora P è vero, allora P è falso, allora P è vero, allora... Le oscillazioni logiche del paradosso sono colte dalla dinamica del valore di verità.»

«E $p=0,5$ è l'unico valore che non porti a un'oscillazione» disse Platone in tono meditativo.

«Esattamente. Ora, il Paradosso duale del biglietto da visita è proprio un caso di dinamica logica:

$$p \leftarrow q$$

$$q \leftarrow 1-p.$$

Supponiamo di iniziare con una stima $p=0,3$, $q=0,8$. La prima correzione dà $p=0,8$, $q=0,7$. Un'ulteriore correzione ci porta a $p=0,7$, $q=0,2$; una terza a $p=0,2$, $q=0,3$. Una quarta correzione dà $p=0,3$, $q=0,8$, e si torna al punto di partenza. È un ciclo di periodo quattro, a meno di partire con $p=0,5$, $q=0,5$: in tal caso, tutto rimane invariato.»

«D'accordo, fin qui mi hai convinto - dissi. - Ma il caos?»

Lukasiewicz si fece serio in viso. «Prima di poter affrontare questo argomento, devo fare una precisazione. Se hai intenzione di continuare a ragionare da solo su questi temi, dovrei dirti come calcolare valori di verità sfumati per combinazioni di enunciati logici (*si veda la finestra in questa pagina*). Per il momento, però, quello che ti serve realmente sapere è che non P ha valore di verità $1-p$ se P ha valore di verità p. In secondo luogo devi sapere come stabilire il valore di verità di enunciati che riguardano enunciati.»

«Vorrei un esempio» disse Socrate.

«D'accordo. Supponiamo che io asserisca che Platone è un buon giocatore di golf. Che valore di verità attribuiresti a questa affermazione?»

«Oh... circa il 40 per cento» disse Socrate. Platone gli lanciò un'occhiata cattiva. «Beh, di solito Epimenide ti batte, e non è certo un gran giocatore.» Questa volta fui io a lanciargli un'occhiata ancor più cattiva.

«Bene. Chiamiamolo enunciato S e diciamo che ha valore di verità $s=0,4$. Supponiamo che io ora produca, a proposito di S, il seguente enunciato T: "S è al 100 per cento vero." Quanto è vero l'enunciato T?»

Logica sfumata

In logica classica un enunciato può assumere solo due valori di verità: 1 se è vero, 0 se è falso. L'enunciato «il Sole splende» ha valore di verità 0 se è nuvoloso. In generale, un enunciato P ha un valore di verità p uguale a 1 o a 0. In logica sfumata, un enunciato può avere un valore di verità compreso tra 0 e 1. Se una nuvola copre un quarto del Sole, l'enunciato P ha valore 0,75.

In logica sfumata, come nella teoria classica, il valore di verità di un enunciato cambia se si applicano gli operatori *non*, *e*, *o*, *implica* e *se e solo se*.

non P ha valore di verità $1-p$.
ESEMPIO: Se «Il Sole splende» ha valore di verità 0,75, allora «*non* splende» ha valore 0,25.

P *e* Q ha valore di verità uguale al più piccolo dei valori p e q, dove q è il valore di verità dell'enunciato Q.
ESEMPIO: P = «Il Sole splende» con $p=0,75$; Q = «Jane si sta abbronzando» con $q=0,10$. «Il Sole splende *e* Jane si sta abbronzando» ha valore 0,10.

P *o* Q ha valore di verità uguale al maggiore dei valori p e q.
ESEMPIO: P = «Il Sole splende» con $p=0,75$; Q = «Jane si sta abbronzando» con $q=0,10$. «Il Sole splende *o* Jane si sta abbronzando» ha valore 0,75.

P *implica* Q ha valore di verità pari al minimo tra 1 e $1-p+q$.
ESEMPIO: P = «Il Sole splende» con $p=0,25$; Q = «Jane si sta abbronzando» con $q=0,10$. «Il Sole splende *implica* che Jane si sta abbronzando» ha valore 0,85.

P *se e solo se* Q ha valore di verità uguale a $1-|p-q|$.
ESEMPIO: P = «Il Sole splende» con $p=0,75$; Q = «Jane si sta abbronzando» con $q=0,10$. «Il Sole splende *se e solo se* Jane si sta abbronzando» ha valore 0,35.

Riflettei un attimo. «Beh, non è certamente vero al 100 per cento; altrimenti *S* sarebbe vero al 100 per cento, e abbiamo già stabilito che non è così.»

«Giusto. Il grado di verità del mio enunciato *T*, che riguarda *S*, dipende dall'effettivo valore di verità di *S* e dal valore di verità attribuito a *S* da *T*. In questo caso, $s = 0{,}4$, ma secondo *T* il valore è 1. Quindi *T* è non vero nella misura in cui i due valori differiscono, no? Più la mia valutazione è imprecisa, più falso diventa il mio enunciato. Dato che la differenza è in questo caso pari a 0,6, *T* è falso nella misura di 0,6. Cioè, è vero nella misura di 0,4.»

«E se tu avessi detto che *S* è vero a metà?»

Lukasiewicz annuì soddisfatto. «Vedrai come funziona bene la cosa. Quell'enunciato stima a 0,5 il valore di verità di *S*, ma il valore effettivo è 0,4. La differenza, 0,1, è la misura della falsità del tuo enunciato: quindi il suo valore di verità è 0,9. Dato che la tua valutazione è errata del 10 per cento, sei nel vero al 90 per cento.»

«Ah. E se avessi detto che *S* è vero al 40 per cento, avrei avuto ragione al 100 per cento: il valore di verità sarebbe stato 1. Ho capito.»

«Ottimo. In generale, supponiamo che io abbia un enunciato *P* con valore di verità *p* e un enunciato *Q* che ci porta a stimare in *p'* il valore di verità di *P*. Allora il ragionamento che abbiamo appena fatto ci dice che il valore di verità di *Q* è $q = 1 - |p - p'|$, dove $|x|$ sta per il valore assoluto di *x* (uguale a *x* quando *x* è positivo, a $-x$ quando *x* è negativo). Chiamiamola formula di stima.»

Lukasiewicz rimase un attimo pensieroso. «Ora posso formulare quello che chiamerò enunciato *C*, o del Mentitore caotico:

Questo enunciato è tanto vero quanto viene stimato falso.

Se il suo valore di verità è *c*, allora esso ci dice di stimare un valore di verità pari a $1 - c$. Secondo la formula di stima, pertanto, il suo valore di verità è $1 - |c - (1 - c)| = 1 - |1 - 2c|$. In breve, abbiamo un processo dinamico

$$c \leftarrow 1 - |1 - 2c|$$

di ricalcolo del valore di verità di *c*. Scegliamo un valore di partenza qualsiasi, per esempio $c = 0{,}12345$, e calcoliamo valori successivi. Scopriremo che sono caotici. In realtà, l'effetto degli errori di arrotondamento fatti da una calcolatrice o da un calcolatore potrebbe dare l'impressione che il processo si stabilizzi su 0 o 1. Forse sarebbe meglio sostituire la precedente formula dinamica con

$$c \leftarrow 1 - |0{,}999999 - 2c|.$$

Potete perfino osservare il famoso effetto farfalla della teoria del caos: il battito d'ali di una farfalla può provocare un uragano un mese dopo. In termini più prosaici, piccole variazioni delle condizioni iniziali producono grandi variazioni nella dinamica successiva. Basta usare 0,12346 come valore di partenza per ottenere un'immagine del tutto diversa.»

Lukasiewicz esitò. «Poi c'è il Duale caotico, che prevede due enunciati:

X : X è tanto vero quanto Y è vero

Y : Y è tanto vero quanto X è falso

È un po' come il tuo biglietto da visita, Epimenide. Le formule dinamiche sono

$$x \leftarrow 1 - |x - y|$$

$$y \leftarrow 1 - |y - (1 - x)|.$$

Per vedere come funziona, scegliamo una coppia iniziale di valori, per esempio $(x, y) = (0{,}2, 0{,}9)$, e calcoliamo coppie successive di valori. Se le tracciamo come coordinate sul piano, otteniamo una figura geometrica chiamata attrattore del sistema dinamico. In questo caso si ottiene un triangolo densamente riempito di punti (*si veda l'illustrazione a destra a pagina 82*). Si può trasformare questa rappresentazione in una bella e complicata immagine chiamata diagramma del tempo di fuga. Per ottenerla rendiamo temporaneamente meno vincolante la condizione che *x* e *y* siano compresi tra 0 e 1. L'idea è quella di osservare quanto (x, y) si allontani dall'origine (0, 0) e di contare quanti passaggi di calcolo siano necessari prima che superi un valore soglia. Si traccia poi il punto (x, y) con un colore che dipende dal numero di passaggi necessari. Per iniziare, si dovrebbe scegliere un valore soglia appena superiore a 1 (*si veda l'illustrazione a sinistra a pagina 82*).

«Comincio a capire» disse Socrate. «Si parte dall'idea del calcolo del valore di verità di un insieme di enunciati autoreferenziali e la si trasforma in un processo dinamico. A questo punto si possono applicare al processo tutte le tecniche della teoria del caos. Il diagramma del tempo di fuga si ispira a un metodo esattamente identico a quello usato per creare le meravigliose immagini associate all'insieme di Mandelbrot: spirali vorticose, cavallucci marini, cactus, stelle e così via.»

«In effetti. Ecco un'ultima idea su cui meditare: si può riformulare a questo modo l'enunciato del Mentitore caotico:

La falsità stimata di questo enunciato non è diversa dalla sua verità stimata.

In logica sfumata, è consuetudine interpretare l'aggettivo "molto" come espressione del quadrato di un valore di verità. Rendiamo allora questo enunciato un po' più vago:

La falsità stimata di questo enunciato non è molto diversa dalla sua verità stimata.

L'enunciato porta alla formula dinamica

$$p \leftarrow 1 - (p - (1 - p))^2$$

che diventa

$$p \leftarrow 4p\,(1 - p).$$

Gli studiosi di teoria del caos chiamano logistico questo sistema dinamico; quindi il mio enunciato è quello del Mentitore logistico. È anch'esso caotico... provate a controllare.»

Verso mezzanotte il Club Paradosso chiudeva. Io e Lukasiewicz uscimmo in strada insieme. Mi resi conto che, tutto assorto negli esempi di logica sfumata caotica, mi ero dimenticato di porre una domanda di grande importanza. «Luke, tutto ciò è molto bello, ma quanto è significativo?»

«Beh, Mar e Grim osservano che fornisce un approccio geometrico alla complessità semantica, permettendo di distinguere tra sistemi differenti di enunciati autoreferenziali. Sostengono anche che può essere usato per dimostrare che non esiste alcuna procedura decisionale che possa dirci se un sistema dato sia caotico oppure no. È un risultato della stessa generalità del famoso teorema di Kurt Gödel sull'indecidibilità dell'aritmetica. Si tratta di roba piuttosto profonda, Epimenide.»

«Vedo. Collegamenti tra logica e caos... stupefacente! Ma aspetta un attimo. Come faccio a essere sicuro che tutto quello che mi hai detto è vero?»

«Che gli dei mi colpiscano con due fulmini se ti ho mai mentito!»

Proprio in quel momento il cielo si riempì di nuvoloni di tempesta e un singolo fulmine bastò a scagliare Lukasiewicz nell'oblio. Alzai gli occhi al cielo, agitando il pugno verso le nubi: «Allora, mi stava dicendo l'intera verità, o soltanto mezza?»

BIBLIOGRAFIA

STEWART IAN, *Does God Play Dice? The Mathematics of Chaos*, Basil Blackwell, 1990.

MAR GARY e GRIM PATRICK, *Pattern and Chaos: New Images in the Semantics of Paradox* in «Noûs», 25, n. 5, dicembre 1991.

GRIM PATRICK, *Self-Reference and Chaos in Fuzzy Logic: Research Report #92-01*. Group for Logic and Formal Semantics, Department of Philosophy, S.U.N.Y. at Stony Brook, 1992.

MAR GARY e GRIM PATRICK, *Computer Investigations in the Semantics of Paradox: Chaotic Liars, Fractals, and Strange Attractors* in «Philosophy and Computing» (in stampa).